ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Προηγμένες Υπηρεσίες Κυβερνοασφάλειας: Χρήση της Τεχνητής Νοημοσύνης για την Ανίχνευση και Αντιμετώπιση Επιθέσεων Ηλεκτρονικού Ψαρέματος και Κοινωνικής Μηχανικής σε Πραγματικό Χρόνο.

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

του

ΚΑΠΑΙ ΑΛΕΞΑΝΤΕΡ

(ΑΕΜ:4069)

Επιβλέπων : ΝΙΚΟΛΑΟΥ ΣΠΥΡΙΔΩΝ

ΛΕΚΤΟΡΑΣ

Καστοριά – Νοέμβριος 2023

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Προηγμένες Υπηρεσίες Κυβερνοασφάλειας: Χρήση της Τεχνητής Νοημοσύνης για την Ανίχνευση και Αντιμετώπιση Επιθέσεων Ηλεκτρονικού Ψαρέματος και Κοινωνικής Μηχανικής σε Πραγματικό Χρόνο.

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

του

ΚΑΠΑΙ ΑΛΕΞΑΝΤΕΡ

(ΑΕΜ:4069)

Επιβλέπων : ΝΙΚΟΛΑΟΥ ΣΠΥΡΙΔΩΝ

ΛΕΚΤΟΡΑΣ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 00 Οκτωβρίου 0000

.....	.....	.....
Ον/μο Μέλους	Ον/μο Μέλους	Ον/μο Μέλους
Ιδίοτητα Μέλους	Ιδίοτητα Μέλους	Ιδίοτητα Μέλους

Καστοριά – Νοέμβριος 2023

Ευχαριστίες

Ευχαριστώ ιδιαιτέρως τον επιβλέποντα καθηγητή μου κύριο Σπυρίδων Νικολάου για τη συνεχή καθοδήγηση και την άψογη συνεργασία σε όλα τα στάδια εκπόνησης της πτυχιακής εργασίας μου.

Επίσης, ευχαριστώ θερμά καθηγητές, γονείς και φίλους που με επηρέασαν θετικά και με υποστήριξαν ενεργά καθ' όλη τη διάρκεια των σπουδών μου στο Τμήμα Πληροφορικής, της Σχολής Θετικών Επιστημών, του Πανεπιστημίου Δυτικής Μακεδονίας.

Περίληψη

Η ραγδαία εξέλιξη του διαδικτύου και του κυβερνοχώρου έχει οδηγήσει την ανθρωπότητα στην ψηφιακή επανάσταση κι έχει επιφέρει επαναστατικές τεχνολογικές εξελίξεις. Χαρακτηριστικά, προσφέρει στους χρήστες απαράμιλλη ευκολία στις διαδικτυακές συναλλαγές, αλλά ταυτόχρονα έχει ανοίξει τις πόρτες σε ποικίλες απειλές ως προς την ασφάλειά τους, μέσω διαφόρων τεχνικών επίθεσης, όπως οι επιθέσεις κοινωνικής μηχανικής και κυρίως το ηλεκτρονικό ψάρεμα (Phishing). Αυτή η ύπουλη μέθοδος, που συνδυάζει την ανθρώπινη χειραγώγηση με τεχνικές παραπλάνησης, έχει ως στόχο να εξαπατήσει τους χρήστες των διαδικτυακών εφαρμογών, ώστε να αποκαλύψουν εμπιστευτικά δεδομένα τους, όπως προσωπικές και οικονομικές πληροφορίες τους. Στο επίκεντρο αυτών των επιθέσεων βρίσκονται παραπλανητικά μηνύματα που συνήθως παραπέμπουν σε κακόβουλες διευθύνσεις (URL), οι οποίες είναι επιδέξια διαμορφωμένες ώστε να μοιάζουν με αυθεντικούς ιστότοπους, καθιστώντας την αναγνώρισή τους εξαιρετικά δύσκολη.

Αναγνωρίζοντας την επείγουσα ανάγκη για βελτιωμένες άμυνες έναντι των επιθέσεων κοινωνικής μηχανικής, πολυάριθμες μελέτες έχουν διερευνήσει προηγούμενες στρατηγικές, επισημαίνοντας τα εμπόδια και προτείνοντας πρωτοποριακά πλαίσια μηχανικής μάθησης. Ειδικότερα, η Βαθιά Εκμάθηση (Deep Learning) και ένα πρωτοποριακό μοντέλο ανίχνευσης αναδεικνύονται ως ελπιδοφόροι διεκδικητές για την ενίσχυση της άμυνας κατά της έξαρσης των επίμονων προσπαθειών ηλεκτρονικού ψαρέματος. Ακόμα και με εκτεταμένη έρευνα σε διάφορες λύσεις μηχανικής μάθησης που μπορούν να χρησιμοποιηθούν από τους αμυνόμενους διαπιστώνεται πως η αποτελεσματικότητά τους παραμένει αμφισβητήσιμη.

Αυτή η πτυχιακή εργασία είναι μια προσπάθεια εισόδου στην πιο σύγχρονη τεχνολογία υπηρεσιών Κυβερνοασφάλειας, με κύριο ζήτημα προσοχής την χρήση της τεχνητής νοημοσύνης για την αναβάθμιση της ασφάλειας τον σημερινών υπολογιστικών συστημάτων. Τα ακόλουθα μοντέλα μηχανικής μάθησης, όπως TCN (Temporal Convolutional Network), LSTM (Long Short-Term Memory) και DNN (Deep Neural Network), έχουν υλοποιηθεί και ερευνηθεί ως προς την αξιοποίησή τους ως προς την ανίχνευση κι αντιμετώπιση των επιθέσεων ηλεκτρονικού ψαρέματος. Στην παρούσα εργασία αναλύονται τα συγκριτικά αποτελέσματα της πειραματικής υλοποίησης των παραπάνω μοντέλων μηχανικής μάθησης και αξιολογείται η απόδοσή τους σε πραγματικό χρόνο, ως προς την ανίχνευση κι αντιμετώπιση των επιθέσεων ηλεκτρονικού ψαρέματος.

Λέξεις Κλειδιά: Κυβερνο-Ασφάλεια, Ευπάθειες, Απειλές, Επιθέσεις, Επιθέσεις Κοινωνικής Μηχανικής, Ηλεκτρονικό Ψάρεμα, Μετριασμός Απειλών, Αντίμετρα, Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Εκπαίδευση Μοντέλων Μηχανικής Μάθησης, TCN, LSTM, DNN

Abstract

The rapid evolution of the internet and cyberspace has led humanity to the digital revolution and has brought about revolutionary technological developments. In particular, it offers users unparalleled convenience in online transactions, but at the same time it has opened the door to a variety of security threats through various attack techniques, such as social engineering attacks and especially phishing. This insidious method, which combines human manipulation with deception techniques, aims to trick users of online applications into revealing confidential data such as personal and financial information. At the heart of these attacks are deceptive messages, usually referring to malicious URLs that are cleverly crafted to look like genuine websites, making them extremely difficult to identify.

Recognising the urgent need for improved defences against social engineering attacks, numerous studies have explored previous strategies, identifying barriers and proposing innovative machine learning frameworks. In particular, Deep Learning and an innovative detection model emerge as promising contenders for enhancing defenses against the surge in persistent phishing attempts. Even with extensive research into various machine learning solutions that can be used by defenders, it is found that their effectiveness remains questionable.

This thesis is an attempt to introduce the latest technology in Cybersecurity services, with the main focus on the use of artificial intelligence to upgrade the security of today's computer systems. The following machine learning models, such as TCN (Temporal Convolutional Network), LSTM (Long Short-Term Memory) and DNN (Deep Neural Network), have been implemented and researched in terms of their utilization towards detecting and countering phishing attacks. In this paper, we analyze the comparative results of the experimental implementation of the above machine learning models and evaluate their real-time performance in detecting and countering the phishing attacks.

Keywords: Cybersecurity, Vulnerabilities, Threats, Attacks, Social Engineering Attacks, Phishing Attacks, Threat Mitigation, Countermeasures, Artificial Intelligence, Machine Learning, Machine Learning Model Training, TCN, LSTM, DNN.

Πίνακας Περιεχομένων

Λίστα Εικόνων

Λίστα Πινάκων

1. Εισαγωγή

Ο κυβερνοχώρος είναι ένα δίκοπο μαχαίρι, ενώ είναι ένας απίστευτος κόσμος που υποστηρίζει την καινοτομία και την εξέλιξη, ταυτόχρονα ανοίγει τις θύρες του σε απειλές Κυβερνοασφάλειας. Καθώς οι οργανισμοί βασίζονται ολοένα και περισσότερο στις ψηφιακές πλατφόρμες, παράλληλα αυξάνεται δραματικά η ανάγκη να λαμβάνονται μέτρα για την ενίσχυση της ασφάλειας. Παρατηρείται σημαντική αύξηση των επιθέσεων ηλεκτρονικού ψαρέματος και γενικότερα επιθέσεων κοινωνικής μηχανικής, χειραγωγώντας την ανθρώπινη ψυχολογία προκειμένου να προβεί σε μοιραία λάθη. Η τεχνητή νοημοσύνη είναι μια πολλά υποσχόμενη τεχνολογία η οποία ήδη παίζει σημαντικό ρολό στον εντοπισμό τέτοιων απειλών κι επιθέσεων κοινωνικής μηχανικής, καθώς και στην απόκρουσή τους σε πραγματικό χρόνο, δίνοντας το παράδειγμα καλής πρακτικής που μπορεί να εφαρμοστεί στην πλειονότητα των υπηρεσιών Κυβερνοασφάλειας.

Η φύση του ηλεκτρονικού ψαρέματος και των επιθέσεων κοινωνικής μηχανικής ελλοχεύουν σοβαρές απειλές και κινδύνους, τόσο σε ομαδικό όσο και σε προσωπικό επίπεδο, ενώ τα παραδοσιακά αντίμετρα Κυβερνοασφάλειας πολλές φορές αποτυγχάνουν να εντοπίσουν εγκαίρως αυτές τις επιθέσεις και να τις αποκρούσουν αποτελεσματικά και σε πραγματικό χρόνο. Υπάρχουν δισεκατομμύρια χρήστες που χρησιμοποιούν το διαδίκτυο καθημερινά.

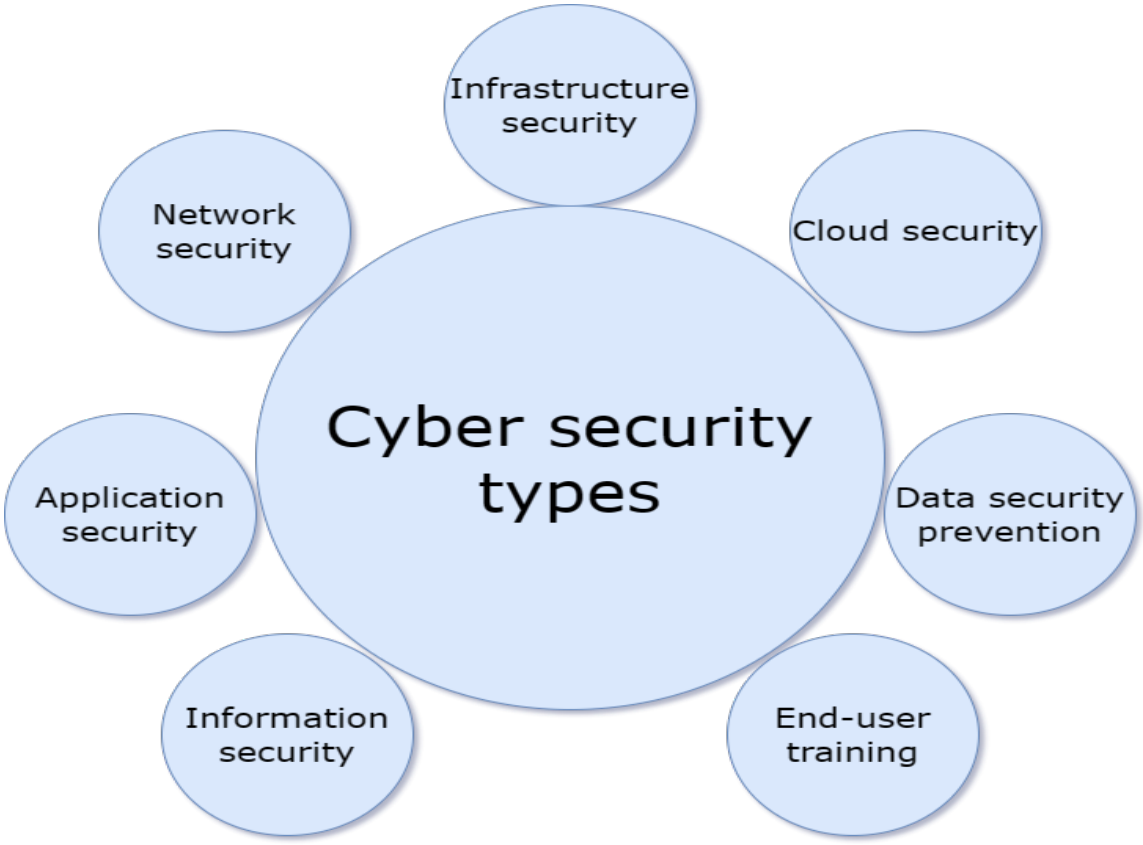
Σκοπός αυτής της έρευνας είναι η μελέτη της αξιοποίησης μοντέλων μηχανικής μάθησης για την αναβάθμιση της ασφάλειας από κακόβουλες επιθέσεις ηλεκτρονικού ψαρέματος και η πειραματική υλοποίηση συστήματος εντοπισμού και αντιμετώπισης επιθέσεων κοινωνικής μηχανικής με βάση μοντέλα μηχανικά μάθησης. Στόχος είναι να διερευνήσουμε σύγχρονες τεχνικές που συνδυάζουν εφαρμογές τεχνικής νοημοσύνης στο περιβάλλον της Κυβερνοασφάλειας, για τον εντοπισμό, σε πραγματικό χρόνο, επιθέσεων κοινωνικής μηχανικής, όπως επίσης να δημιουργήσουμε, να παραμετροποιήσουμε και να αξιολογήσουμε τρία μοντέλα τεχνητής νοημοσύνης (Temporary Convolutional Network, Deep Neural Network, Long Short-Term Memory), ως προς τη λειτουργικότητα και την αποδοτικότητα τους, σε ένα πραγματικό σκηνικό επιθέσεων ηλεκτρονικού ψαρέματος. Το πειραματικό μέρος της εργασίας εστιάζεται συγκεκριμένα στην υλοποίηση των μοντέλων μηχανικής μάθησης που προαναφέρθηκαν, με έμφαση στο ηλεκτρονικό ψάρεμα.

Η έκταση της έρευνας και των πειραμάτων που περιλαμβάνονται στην παρούσα εργασία είναι περιορισμένη λόγω της έλλειψης επαρκών δεδομένων επιθέσεων ηλεκτρονικού ψαρέματος, αλλά και διαθέσιμων υπολογιστικών πόρων.

2. Υπηρεσίες Κυβερνοασφάλειας

2.1 Εισαγωγή στην Κυβερνοασφάλεια

Η Κυβερνοασφάλεια αποτελεί ένα σημαντικό παράγοντα στη σημερινή πραγματικότητα σε κάθε οργανισμό ή εταιρεία, αλλά και σε μεμονωμένους χρήστες. Με λίγα λόγια ένας οργανισμός ή μια εταιρεία που επενδύει στην ασφάλειά της όσων αφορά τον κυβερνοχώρο επιτυγχάνει ένα επίπεδο κύρους και εμπνέει εμπιστοσύνη, κάτι που ύστερα φέρνει την επιτυχία. Αυτό συμβαίνει επειδή η επιτυχία αυτή βασίζεται στην τεχνογνωσία και ικανότητα προστασίας δεδομένων έναντι κάποιου ανταγωνιστή. Οι επιστήμονες της Κυβερνοασφάλειας προστατεύουν υπολογιστικά συστήματα, δίκτυα, διακομιστές και γενικότερα κάθε είδους συσκευών. Για παράδειγμα, διασφαλίζουν πως μόνο εξουσιοδοτημένοι χρήστες θα μπορούν να έχουν πρόσβαση σε κρίσιμες υπηρεσίες και ευαίσθητα δεδομένα [1]. Η εικόνα 1 περιλαμβάνεις τις βασικές κατηγορίες/τύπους της Κυβερνοασφάλειας [2] [3] [4] [5].



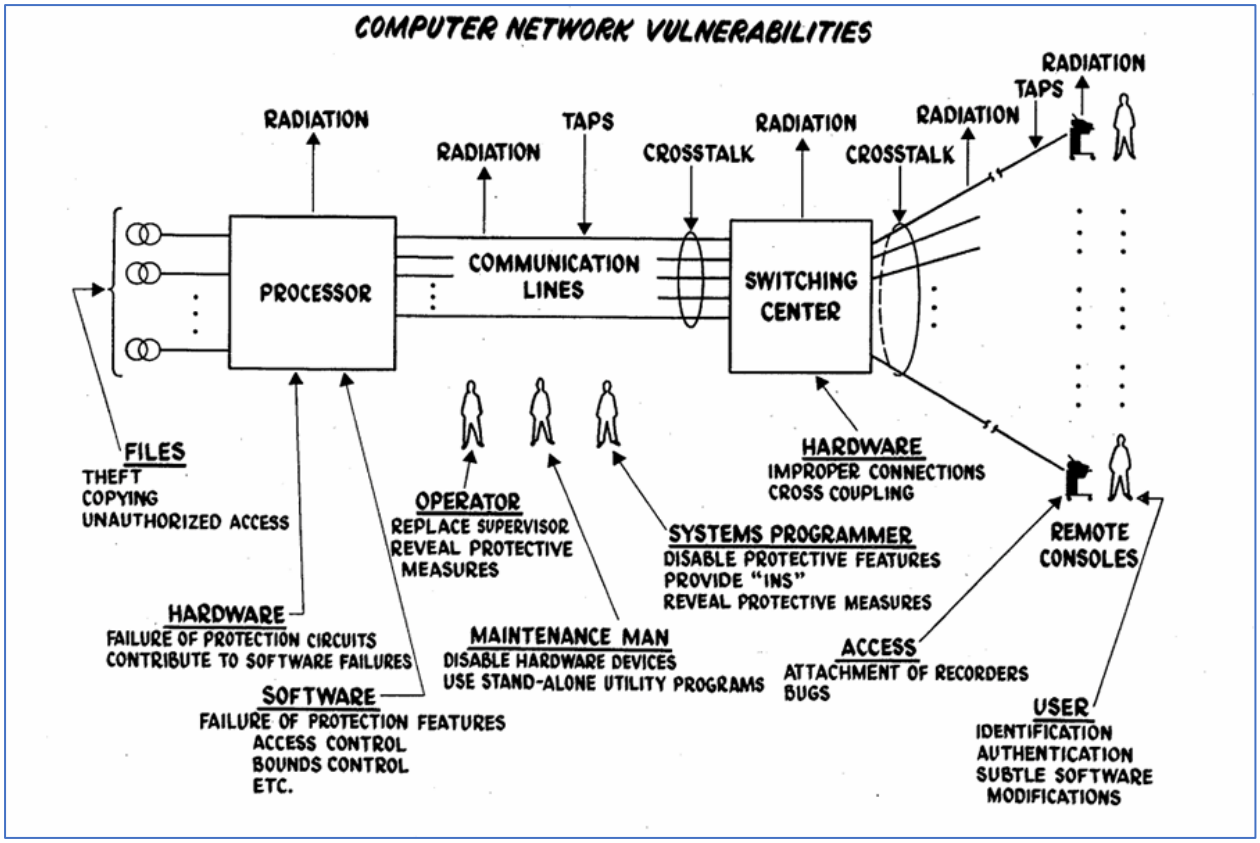
Εικόνα 1. Τύποι Κυβερνοασφάλειας

- **Application Security:** Στην ασφάλεια εφαρμογών χρησιμοποιώντας κατάλληλο hardware και software, όπως είναι τα τείχη προστασίας και η κρυπτογράφηση, προστατεύουν τα πληροφοριακά συστήματα και οι εφαρμογές τους από κακόβουλες ενέργειες.
- **Information Security:** Στην ασφάλεια πληροφοριών παρέχεται προστασία φυσικών και ψηφιακών πληροφοριών ενάντια σε μη εξουσιοδοτημένη πρόσβαση, απόκρυψη, κατάχρηση, αλλαγή και διαγραφή.
- **Operational Security:** Στη λειτουργική ασφάλεια περιέχονται διαδικασίες και αποφάσεις που αφορούν την πιστοποίηση ταυτότητας, την εξουσιοδότηση και την καταγραφή όλων των συμβάντων που αφορούν τη λειτουργία ενός συστήματος.
- **Cloud Security:** Στην ασφάλεια του υπολογιστικού νέφους περιλαμβάνεται η προστασία των πληροφοριών που βρίσκονται σε απομακρυσμένα κέντρα δεδομένων υπολογιστικού νέφους.
- **User Training:** Η εκπαίδευση χρηστών αναφέρεται στον πιο αδύναμο κρίκο της Κυβερνοασφάλειας που είναι οι ίδιοι οι χρήστες των πληροφοριακών συστημάτων κι εφαρμογών. Όλοι μας μπορεί κατά λάθος να κολλήσουμε κάποιον ιό στον υπολογιστή μας ή να το διαπεράσουμε από το σύστημα ασφάλειας. Έτσι εκπαιδεύοντας τον χρήστη να αφαιρεί ύποπτες επισυνάψεις σε μια ηλεκτρονική αλληλογραφία και γενικότερα άλλα σοβαρά λάθη που μπορούν να συμβούν.
- **Network Security:** Η ασφάλεια δικτύων διαδραματίζει ένα σημαντικό ρόλο ως προς την Κυβερνοασφάλεια. Σε γενικές γραμμές, αφορά την προστασία δικτύων υπολογιστών από μη εξουσιοδοτημένη πρόσβαση, υποκλοπή ή άλλη κακόβουλη ενέργεια. Περιλαμβάνει τη δημιουργία ασφαλούς υποδομής για δικτυακές συσκευές, διαδικτυακές εφαρμογές και χρήστες τους.

2.2 Ιστορική εξέλιξη της Κυβερνοασφάλειας

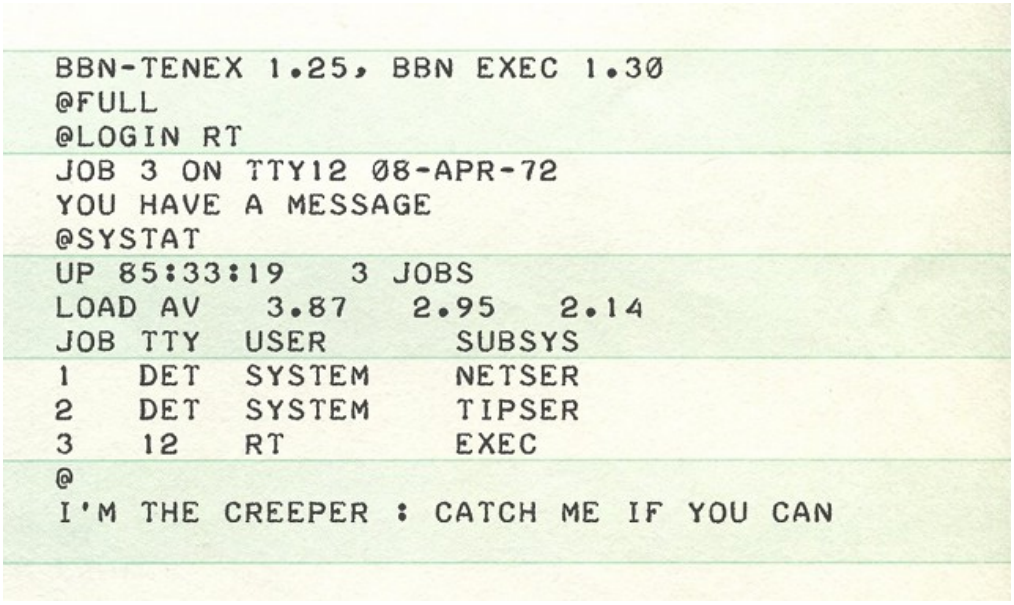
Ως αφετηρία της έννοιας της Κυβερνοασφάλειας μπορεί να θεωρηθεί στο όχι και τόσο μακρινό 1970, με τη δημοσίευση της γνωστής έρευνας “Security Controls of Computer Systems”, επίσης γνωστή και ως WARE Report [6]. Αυτό το άρθρο προέκυψε από τη σχετική έρευνα την οποία διεξήγαγε το Υπουργείο Άμυνας των Ηνωμένων Πολιτειών της Αμερικής, με κύρια έμφαση την ανάγκη για μια καθολική ασφάλεια που αφορούσε υλισμικό, λογισμικό, επικοινωνίες και διάφορες άλλες τεχνολογίες.

Η αυγή του 1970 επίσης είδε την δημιουργία του πρώτου υπολογιστικού προγράμματος με την δυνατότητα πλοήγησης στο δίκτυο, το όνομά του ήταν Creeper, που άφηνε πίσω του ένα μονοπάτι μηνυμάτων οπουδήποτε και αν πήγαινε. Με στόχο την μάχη ενάντια σε αυτόν τον ιό, το πρώτο παράδειγμα ενός αντιικού προγράμματος με όνομα Reaper δημιουργήθηκε το 1973, το οποίο κυνήγησε και εξαφάνισε τον ιό Creeper.



Εικόνα 2. Ευπάθειες των Δικτύων Υπολογιστών σύμφωνα με το Ware Report

Πηγή: https://www.digitaltrends.com/computing/cyberscurity-1970-memo-ware-report/



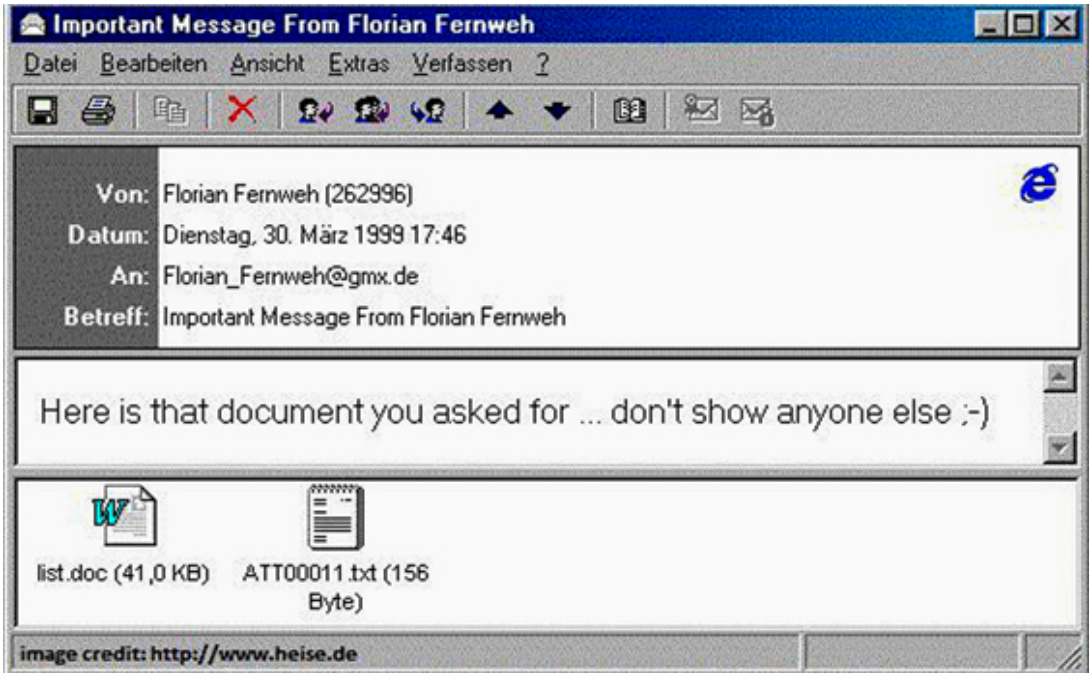
Εικόνα 3. Μήνυμα που άφηνε το Creeper

Πηγή: https://isc.sans.edu/diary/rss/27208

Μεταβαίνοντας στο 1977, καθιερώθηκε η τριάδα των θεμελιωδών αρχών της κυβερνοασφάλειας, γνωστή και ως CIA: Εμπιστευτικότητα (Confidentiality), Ακεραιότητα (Integrity), Διαθεσιμότητα (Availability). Αν και δεν αποτελεί κάποια μεμονωμένη τεχνολογία, ακόμα και έτσι χρησίμευσε ως ένα σκαλοπάτι για ανερχόμενα πλαίσια και πολιτικές ασφαλείας χρηστών.

Το 1980 ήταν ο χρόνος που ένα σημαντικό κατόρθωμα εμφανίστηκε με την μεταμόρφωση του ARPANET στο INTEPNET, που παραδόξως έγινε αφορμή για την ανάπτυξη κι εξάπλωση νέων και πιο ανεπτυγμένων υπολογιστικών ιών. Ένας αξιοσημείωτος ιός από αυτούς ήταν το σκουλήκι Morris το 1988 το οποίο επίσης βοήθησε στην δημιουργία ακόμα πιο ανεπτυγμένων σκουληκιών (Worms) και ιών βάζοντας τον κόσμο σε μια έκτακτη ανάγκη για κάποιες αντιικές λύσεις.

Στην δεκαετία του 1990, ο χώρος της Κυβερνοασφάλειας εμφάνισε μια σημαντική αύξηση των επιθέσεων από κακόβουλα λογισμικά, ένα από αυτά ήταν ο ιός Melissa, που προκάλεσε μεγάλη αναστάτωση παγκοσμίως, με ραγδαία εξάπλωση μέσα σε λίγες ώρες, εκμεταλλευόμενος ευπάθεια στα συστήματα ηλεκτρονικής αλληλογραφίας. Όμως παράλληλα ξεπρόβαλλε ένας μεγάλος αριθμός από εταιρείες που είχαν ως ειδικότητα τον εντοπισμό και την αντιμετώπιση ιών, σκουληκιών και λοιπόν κακόβουλων λογισμικών. Όμως πολλά από αυτά είχαν προβλήματα με ψευδείς συναγερμούς, κάτι το οποίο μέχρι και σήμερα αποτελεί σημαντικό πρόβλημα αξιοπιστίας. Ένα άλλο πρόβλημα, το οποίο υπήρχε με αυτά τα συστήματα antivirus ήταν η απαίτηση σε μεγάλη χρήση των υπολογιστικών πόρων για την καθημερινή λειτουργία τους.



Εικόνα 4. Melissa virus email

Πηγή: https://www.techpout.com/cyber-threat-latest-computer-viruses-and-malware/

Στη δεκαετία του 2000 οι τεχνολογικές εξελίξεις στο χώρο των υπολογιστικών συστημάτων, συνδυάστηκαν με την εμφάνιση καινούριων απειλών από διάφορους τύπους κακόβουλων λογισμικών, αλλά και καλύτερες μεθόδους προστασίας έναντι αυτών. Μερικές από αυτές τις απειλές ήταν πολυμορφικά και μεταμορφωτικά κακόβουλα λογισμικά, μεταφέροντας τις απειλές τους σε άλλες συσκευές, όπως έξυπνες κινητές συσκευές. Αργότερα εκείνη την δεκαετία ξεπρόβαλε και το πρώτο κακόβουλο λογισμικό που χρησιμοποιήθηκε ως ψηφιακό όπλο κυβερνοπολέμου, το Stuxnet worm, με αποτέλεσμα να δημιουργήσει σοβαρή ζημιά στο ερευνητικό κέντρο πυρηνικών επιστημών στην περιοχή Natanz του Ιράν, προκαλώντας σημαντική χρονική καθυστέρηση στο πυρηνικό πρόγραμμα του Ιράν. Λέγεται ότι το Stuxnet μπορεί να είχε φτιαχτεί από το 2005 από τις μυστικές υπηρεσίες του Ισραήλ ή και των ΗΠΑ και ας ανακαλύφθηκε το 2010.

Εικόνα 5. Το κέντρο πυρηνικών δοκιμών του Ιράν που επλήγη από το Stuxnet worm

Πηγή: https://www.wired.com/2014/11/countdown-to-zero-day-stuxnet/

Τη δεκαετία του 2010, οι τεχνολογικές εξελίξεις ιδίως στο χώρο του διαδικτύου των πραγμάτων (Internet of Things – IoT), συνδυάστηκαν με την εμφάνιση καινούριων απειλών από διάφορους τύπους κακόβουλων λογισμικών. Επίσης, τα τελευταία έτη, έχει παρατηρηθεί μεγάλη έξαρση σε κακόβουλες επιθέσεις τύπου Ransomware, όπως για παράδειγμα το πασίγνωστο WannaCry, αλλά και το Clop Ransomware και το Mount Locker.

Τη δεκαετία του 2020, η χρήση της τεχνητής νοημοσύνης έχει διεισδύσει σε όλους τους τομείς των τεχνολογιών πληροφορίας κι επικοινωνιών, ωστόσο έχει αρχίσει να χρησιμοποιείται ευρέως και από τους επίδοξους κυβερνοεγκληματίες, οι οποίοι αναβαθμίζουν τις τακτικές τους, εφαρμόζοντας προηγμένες επίμονες απειλές. Στο μοντέρνο αυτό κυβερνοπόλεμο, οι συνεχείς εξελίξεις στον τομέα της Κυβερνοασφάλειας αναγκάζει και καταδικάζει και τις δύο πλευρές, επιτιθέμενους κι αμυνόμενους, σε έναν μόνιμο αγώνα δρόμου, πιθανότατα χωρίς τερματισμό.

Εικόνα 6. Screenshot of the WannaCry ransom note left on an infected system

Πηγή: https://www.d4f.co.za/

2.3 Κυβερνοασφάλεια και έξυπνες κινητές συσκευές

Ας κάνουμε μια μικρή επισκόπηση των διάφορων προηγμένων τεχνικών Κυβερνοασφάλειας, συμπεριλαμβάνοντας τον εντοπισμό και την αξιολόγηση ευπαθειών, την παρακολούθηση και διαχείριση κακόβουλων ενεργειών και την αντιμετώπισή τους, τόσο σε κλασικά υπολογιστικά συστήματα, όσο και στις έξυπνες φορητές συσκευές. Τεράστιο κομμάτι της σημερινής κοινωνίας χρησιμοποιεί διαρκώς τα έξυπνα κινητά τηλέφωνα και τις εφαρμογές των έξυπνων φορητών συσκευές, από μια απλή φωτογραφία, μέσα κοινωνικής δικτύωσης, έως και τραπεζικές συναλλαγές. Φυσικά, αυτή η εξέλιξη προκάλεσε γρήγορα το ενδιαφέρον των επίδοξων επιτιθέμενων και αποτέλεσε πεδίο εξαπόλυσης επιθέσεων. Κατά την τελευταία δεκαετία έχουν πολλαπλασιαστεί εκθετικά οι κακόβουλες επιθέσεις διαφόρων τύπων, φτάνοντας σε πλήθος δισεκατομμυρίων επιθέσεων παγκοσμίως. Αντίστοιχα, έχει γίνει πιο σύνθετο και το έργο αντιμετώπισής τους.

Οι ερευνητές έχουν χωρίσει την απόκρουση των επιθέσεων σε δύο κατηγορίες [8].

- **Signature based**: Αυτές οι υπηρεσίες βασίζονται σε υπογραφές συγκεκριμένων μοτίβων (patterns) κακόβουλων λογισμικών, για έγκαιρο εντοπισμό τους και τη λήψη αντιμέτρων. Ας εξηγήσουμε λίγο το πως λειτουργούν αυτές οι υπηρεσίες. Σε

συνδυασμό με μια ειδική βάση δεδομένων που διατηρεί το λογισμικό προστασίας από κακόβουλο λογισμικό, όταν σαρώνουν έναν υπολογιστή ή μια συσκευή γενικότερα, θα ψάξουν ίχνη (υπογραφές) που μοιάζουν με αυτές γνωστών κακόβουλων λογισμικών, που περιέχονται ήδη στη βάση δεδομένων. Εάν βρουν κάτι τέτοιο, τότε θα το αναγνωρίσουν ως κακόβουλο λογισμικό και είτε θα το διαγράψουν, είτε θα το τοποθετήσουν σε καραντίνα. Αποτελεί μια πολύ αποτελεσματική τεχνική, όμως υπάρχει τρόπος αντιμετώπισης από τους επιτιθέμενους, κάνοντας πολύ μικρές αλλαγές στην καινούρια έκδοση της εφαρμογής. Για παράδειγμα, ένα τέτοιο εργαλείο ανάλυσης ονομαζόμενο LimonDroid [7] έχει δημιουργηθεί για να εντοπίζει επιβλαβή χαρακτηριστικά σε εφαρμογές κινητών τηλεφώνων. Τα βήματα που ακολουθούνται είναι αρχικά ο εντοπισμός ενός καινούριου κακόβουλου λογισμικού, έπειτα η σύγκριση της υπογραφής του με τις αποθηκευμένες υπογραφές στην ειδική βάση δεδομένων (απαιτείται τακτική ενημέρωση της βάσης δεδομένων), έτσι ώστε στο τελευταίο στάδιο να μπορεί να εντοπίσει το ίδιο κακόβουλο λογισμικό εάν εμφανιστεί ξανά.

- **Behavior Based**: Οι υπηρεσίες που βασίζονται στη διερεύνηση της συμπεριφοράς των εφαρμογών και των χρηστών, χρησιμοποιούνται σε μεγάλο βαθμό από τα σύγχρονα εργαλεία προστασίας έναντι των κακόβουλων λογισμικών. Αυτή η τακτική μπορεί να χρησιμοποιηθεί για να αναγνωρίσει γνωστούς και άγνωστους κινδύνους από κακόβουλα λογισμικά. Πολλοί ερευνητές και πάροχοι υπηρεσιών ασφαλείας προτιμούν τεχνικές αυτού του τύπου, οι οποίες συνήθως χρησιμοποιούν την τεχνητή νοημοσύνη. Ένας κύριος λόγος που συμβαίνει αυτό είναι το γεγονός ότι είναι πιο αποτελεσματικές ως προς τον εντοπισμό προηγμένων απειλών, όπως Zero-day attacks. Στην έρευνα [9], παρουσιάζεται η αποδοτικότητα των αλγορίθμων μηχανικής μάθησης ως προς την βαθμονομημένη αξιολόγηση των επικίνδυνων δεδομένων σε εφαρμογές έξυπνων κινητών συσκευών, εντοπίζοντας εξαιρετικά ανώμαλα μοτίβα κίνησης δικτύου. Δεδομένα σε δίκτυα κινητών επικοινωνιών, τα οποία έχουν εκτεθεί σε κακόβουλα λογισμικά, είναι από τα πιο ενδεικτικά παραδείγματα ασύμμετρων συνόλων δεδομένων. Ωστόσο, σε πραγματικές συνθήκες καμία τακτική εντοπισμού κι αποτροπής επιθέσεων δεν μπορεί να είναι 100% αποτελεσματική. Εννοείται πως αυτές οι υπηρεσίες δεν είναι περιορισμένες μόνο σε εφαρμογές έξυπνων κινητών συσκευών, αλλά γενικότερα σε όλες τις υπολογιστικές συσκευές που γνωρίζουμε και μπορούν να συνδεθούν στο διαδίκτυο και διαθέτουν λογισμικό προστασίας από κακόβουλο λογισμικό.

2.4 Προηγμένες Υπηρεσίες Κυβερνοασφάλειας

Η κυβερνοασφάλεια λειτουργεί σαν μια ασπίδα που με αξιόλογες προσπάθειες προστατεύει την κοινωνία και την οικονομία από ψηφιακές απειλές που θα μπορούσαν να γκρεμίσουν τις θεμελιώδεις αρχές της Εμπιστευτικότητας, της Ακεραιότητας και της Διαθεσιμότητας, σε ένα σύνολο τομέων της καθημερινότητας, συμπεριλαμβανομένων την

Επιχειρηματικότητα, την Κυβέρνηση, τις Στρατιωτικές δυνάμεις, την Υγεία, την Εκπαίδευση και την Ενέργεια [10]..

Το σύγχρονο πεδίο της κυβερνοασφάλειας χαρακτηρίζεται από τη συνεχή εμφάνιση καινούριων προηγμένων απειλών. Αυτές οι απειλές που γίνονται ολοένα και περισσότερο πολύπλοκες και μοναδικές, αποτελούν σοβαρούς παράγοντες κινδύνων τόσο σε ιδιώτες όσο και σε οργανισμούς. Χαρακτηριστικά, το 2022, οι σημαντικότερες απειλές που συνιστούν μέγιστο βαθμό κινδύνου, ήταν μια μεγάλη γκάμα από επιθέσεις ηλεκτρονικού ψαρέματος έως και κατανεμημένες επιθέσεις άρνησης εξυπηρέτησης (DDoS). Αυτές και άλλες εξαιρετικά επικίνδυνες επιθέσεις μπορούν να είναι λόγος για σημαντικές ζημιές, όπως για παράδειγμα οικονομικές επιπτώσεις, δυσφήμηση, νομικά προβλήματα, ακόμα και απειλή στην εθνική ασφάλεια [11]. Ενώ πολλοί τομείς επενδύουν στην ψηφιακή τους ασφάλεια, άλλοι τομείς μπορεί να μην έχουν τέτοια δυνατότητα, με αποτέλεσμα να δυσχεραίνεται σε μεγάλο βαθμό η ομαλή λειτουργεία τους.

Με σκοπό να αναβαθμίσουν τις άμυνες τους, χρησιμοποιούν μια μεγάλη γκάμα από εργαλεία και μεθοδολογίες που περιλαμβάνουν συστήματα εντοπισμού κι αποτροπής εισβολών IDS/IPS, τείχη προστασίας (firewalls), λύσεις ενάντια σε κακόβουλο λογισμικό και πάνω από όλα κρυπτογράφηση δεδομένων [12]. Επιπροσθέτως υπάρχει μια γενική εστίαση στις καλές συνήθειες του χώρου, όπως είναι οι τακτικές αναβαθμίσεις , καλύτεροι κωδικοί, κ.α. Σε μια προσπάθεια να αναβαθμιστεί η κυβερνοασφάλεια και να προστατευτεί η ατομική ιδιωτικότητα, πολλές χώρες σε Αμερική και η Ευρώπη έχουν θέσει κάποιους θεσμούς για την κοινωνική και συλλογική ασφάλεια [13]. Μέχρι πρότινος άγνωστες τεχνολογίες, όπως το blockchain, αξιοποιούνται πλέον για την ενίσχυση της ασφάλειας του τεράστιου όγκου δεδομένων (Big Data) που διακινούνται παγκοσμίως, άλλες τεχνολογίες όπως η τεχνητή νοημοσύνη, η μηχανική μάθηση και η βαθιά μάθηση αποφέρουν καρπούς για την καλύτερη αντιμετώπιση των ολοένα πιο περίπλοκων απειλών που προκύπτουν καθημερινά.

2.4.1 Κυβερνοασφάλεια στο Smart Grid

Είναι ένα ανεπτυγμένο έξυπνο ηλεκτρικό δίκτυο νέας γενιάς, το οποίο, συνδυάζοντας τεχνολογίες πληροφορίας και επικοινωνιών (ΤΠΕ) με την τεχνητή νοημοσύνη και τις ανανεώσιμες πηγές ενέργειας, ενισχύει την αποδοτικότητα και λειτουργικότητα των συστημάτων ενέργειας του αύριο. Το τωρινό ηλεκτρικό δίκτυο είχε δημιουργηθεί πολλά χρόνια πριν, όταν ήταν πολύ πιο απλές οι ανάγκες παροχής ηλεκτρικής ενέργειας στην κοινωνία. Το smart grid παρέχει συνδυασμό δικτύων παροχής ηλεκτρικής ενέργειας και δικτύων δεδομένων μεταξύ παρόχων ενέργειας και των πελατών τους. Με τον χρόνο ανταπόκρισης να μειώνεται συνεχώς το Smart Grid προσφέρει γρήγορες και κυρίως προηγμένες υπηρεσίες [12].

Τα ζητήματα της κυβερνοασφάλειας που αφορούν στο Smart Grid αποτελούν ένα πολύ κρίσιμο θέμα, μιας και τα συστήματα παροχής ενέργειας έχουν ζωτική σημασία για την

ευημερία της ανθρωπότητας και ενδεχόμενη αποτυχία σε θέματα ασφάλειας θα μπορούσε να προκαλέσει διακοπές ρεύματος, οικονομικές απώλειες και πολύ σοβαρούς κινδύνους για την κοινωνία. Όπως και στον υπόλοιπο χώρο της κυβερνοασφάλειας, έτσι και εδώ πρέπει να διασφαλιστούν οι θεμελιώδεις αρχές, όπως η εμπιστευτικότητα, η ακεραιότητα και η διαθεσιμότητα (CIA). Αυτό επιτυγχάνεται με τη χρήση τεχνικών ταυτοποίησης έτσι ώστε να περιοριστεί στο ελάχιστο η μη εξουσιοδοτημένη πρόσβαση στο Smart Grid, με εφαρμογή πρωτοκόλλων ασφαλούς επικοινωνίας, κρυπτογράφηση των δεδομένων και συνεχή επίβλεψη του δικτύου για εντοπισμό περίεργων μοτίβων κίνησης και άμεση λήψη αντιμέτρων. Εφαρμόζοντας λοιπόν, μηχανισμούς ασφαλείας με βάση τις αρχές CIA, το Smart Grid μπορεί να αποκρούει πιθανούς κινδύνους και ποικίλες απειλές, όπως για παράδειγμα επιθέσεις DDoS, code injections, κλπ. Σύμφωνα με το NIST (National Institute of Standards and Technology) [14] θα πρέπει να δίνεται έμφαση στην εφαρμογή μηχανισμών κρυπτογραφίας για microgrids που συνθέτουν το Smart Grid. Αυτή η καινούρια τεχνική ιδέα έχει σκοπό την προσαρμογή των υφιστάμενων μεθοδολογιών και εργαλείων βέλτιστων πρακτικών κυβερνοασφάλειας με μοναδικό επιθυμητό αποτέλεσμα την εφαρμογή τους στον τομέα παροχής ενέργειας.

2.4.2 Κυβερνοασφάλεια σε έξυπνα οχηματικά δίκτυα κι έξυπνες μεταφορές

Οι προηγμένες τεχνολογίες που βοηθάνε τον οδηγό κατά τη διάρκεια της οδήγησης στα σύγχρονα οχήματα, χρησιμοποιούν μια σειρά από μικροαισθητήρες, ηλεκτρονικά και μικροϋπολογιστικά συστήματα με σκοπό την καλύτερη εμπειρία οδήγησης και την ασφάλεια. Κατά την προώθηση αυτών των χαρακτηριστικών σε μια καινούρια τεχνολογία οχημάτων, η αυτοκινητοβιομηχανία επικεντρώνεται στην έρευνα για θέματα ασφαλείας που αφορούν τα έξυπνα οχηματικά δίκτυα και τις έξυπνες μεταφορές, με την καλύτερη δυνατή αξιοποίηση των διαθέσιμων εργαλείων της Κυβερνοασφάλειας, για να διασφαλίσει ότι αυτά τα συστήματα θα λειτουργούν απρόσκοπτα κι επιπλέον θα περιορίζουν τους όποιους κινδύνους [15].

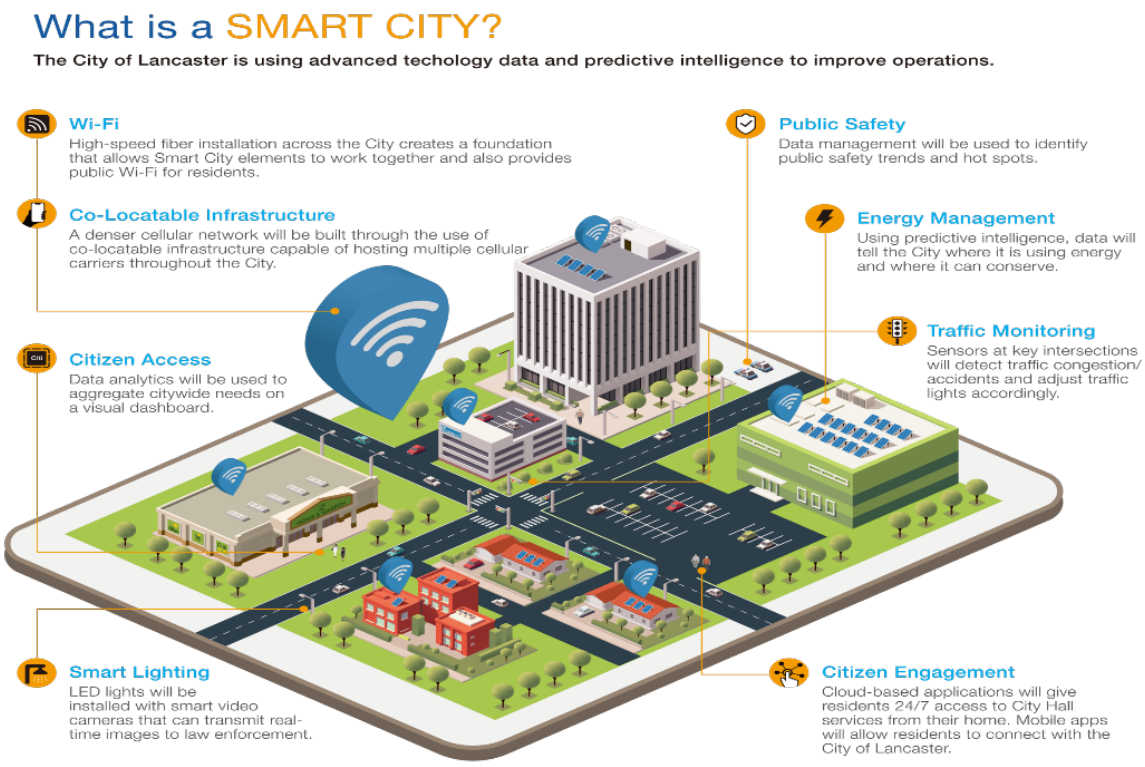
Ο οργανισμός NHTSA (National Highway Traffic Safety Administration) προωθεί μια προσέγγιση με πολλά επίπεδα, εστιάζοντας στα σημεία εισόδου των οχηματικών δικτύων, είτε είναι ασύρματα, είτε ενσύρματα, τα οποία μπορεί να έχουν κενά ασφαλείας που ενδεχομένως να χρησιμοποιηθούν για μια κακόβουλη επίθεση. Μια προσέγγιση, η οποία περιλαμβάνει πολλαπλά επίπεδα ασφάλειας σε οχηματικά δίκτυα, μειώνει την πιθανότητα μιας επιτυχούς επίθεσης και περιορίζει τους όποιους κινδύνους, μέσω των εξής:

- Λειτουργία βασισμένη στη μείωση του ρίσκου, που βάζει σε προτεραιότητα την αναγνώριση και την προστασία των κρίσιμων συστημάτων ελέγχου για την ασφάλεια οχημάτων.
- Διαχρονικός εντοπισμός και ταχεία ανταπόκριση σε πιθανά θέματα κυβερνοασφάλειας των οχηματικών δικτύων.
- Αρχιτεκτονικές, μέθοδοι και μέτρα πρόληψης, ειδικά σχεδιασμένα και εγκατεστημένα για γρήγορη ανάκαμψη από κακόβουλες επιθέσεις.

- Διαμοιρασμός πληροφοριών και τεχνογνωσίας για θέματα ασφαλείας των οχηματικών δικτύων σε βιομηχανικό εύρος.

2.4.3 Κυβερνοασφάλεια σε Smart Cities

Ο όρος έξυπνες πόλεις αναφέρεται σε κοινότητες οι οποίες εφαρμόζουν πληροφοριακές και επικοινωνιακές τεχνολογίες (ΤΠΕ), δεδομένα κοινωνικού εύρους και ευφυής λύσεις για να μεταμορφώσουν ψηφιακά τις υποδομές και φέρουν την διακυβέρνηση στα μέτρα των πολιτών. Αναφέρεται επίσης στη διασύνδεση της επιχειρησιακής τεχνολογίας που διαχειρίζεται υλικές υποδομές, δίκτυα κι εφαρμογές που συλλέγουν και αναλύουν δεδομένα, χρησιμοποιώντας πληροφοριακές και επικοινωνιακές τεχνολογίες, όπως το διαδίκτυο των πραγμάτων, υπολογιστική νέφους, τεχνητή νοημοσύνη και δίκτυα νέας γενιάς. Άλλες γνωστές ορολογίες για τις έξυπνες πόλεις είναι συνδεδεμένα μέρη, συνδεδεμένες κοινωνίες και έξυπνα μέρη.



Εικόνα 7. Έξυπνη πόλη

Πηγή: https://www.cityoflancasterca.org/our-city/departments-services/development-services/city-engineering/traffic-engineering/advanced-traffic-management-system

Οι έξυπνες εφαρμογές που υιοθετούνται από τις έξυπνες πόλεις αφορούν ποικίλα πεδία, όπως για παράδειγμα τη διαχείριση απορριμμάτων, τη διαχείριση του δημοτικού φωτισμού, της οδικής σήμανσης και αστικής κυκλοφορίας, των θέσεων στάθμευσης, κλπ. Το να εφαρμοστούν έξυπνες υπηρεσίες σε ένα συνδεδεμένο περιβάλλον μπορεί να αυξήσει την αποδοτικότητα και την ανθεκτικότητα μιας υποδομής που υποστηρίζει την καθημερινή ζωή των κατοίκων μιας έξυπνης πόλης. Όμως κοινότητες που θέλουν να υιοθετήσουν αυτές τις έξυπνες εφαρμογές, πρέπει να έχουν επίγνωση των απειλών και κινδύνων ασφαλείας και να αναλύουν το ρίσκο που περιλαμβάνεται σε αυτές. Οι έξυπνες πόλεις είναι εξωπραγματικά ελκυστικές στους εγκληματίες του κυβερνοχώρου, προφανέστατα επειδή οι πληροφορίες που μεταφέρονται, αποθηκεύονται και επεξεργάζονται μέσα στην κοινότητα, μπορεί να περιέχουν ευαίσθητα δεδομένα από κυβερνήσεις, εταιρίες ή ιδιώτες. Άλλος ένας λόγος είναι τα ευάλωτα σημεία που υπάρχουν μέσα στα πολύπλοκα συστήματα λογισμικού των έξυπνων εφαρμογών τους. Τα ρίσκα είναι πολλά, αλλά παρακάτω θα δούμε πολλές αποτελεσματικές τεχνικές [16] για να προστατεύσουμε μια έξυπνη πόλη.

- **<u>Ασφαλής προ-σχεδίαση</u>**: Προτείνεται να υπάρχει μια πρόβλεψη για μηχανισμούς ασφαλείας όταν σχεδιάζεται και χτίζεται μια έξυπνη πόλη. Αυτό συμβαίνει επειδή κάθε φορά που θα χρειαστεί μια αναβάθμιση πρέπει ο σχεδιασμός να επιτρέπει την πρόσβαση και εύκολη αλλαγή του παλιού εξοπλισμού με τον καινούριο. Επιπλέον θα πρέπει κατά το σχεδιασμό των συστημάτων της έξυπνης πόλης να καθορίζεται όχι μόνο ασφάλεια των δεδομένων και πληροφοριών, αλλά και ο συνδυασμός ασφάλειας τόσο για το υλισμικό, όσο και για το λογισμικό του οικοσυστήματος που προσφέρει το διαδίκτυο των πραγμάτων.
- **<u>Λίγα προνόμια</u>**: Οι υπεύθυνο οργανισμοί που εφαρμόζουν τις τεχνολογικές λύσεις των έξυπνων πόλεων, θα πρέπει να εφαρμόσουν τον κανόνα των ελάχιστων προνομίων σε όλο το οικοσύστημα της έξυπνης πόλης. Το NIST δηλώνει πως μια υποδομή ασφάλειας πρέπει να σχεδιαστεί έτσι ώστε κάθε οντότητα να έχει τόσα προνόμια όσα απαιτούνται για την ορθή λειτουργία της, τίποτα παραπάνω. Επίσης οι διαχειριστές πρέπει να ελέγχουν ότι όλα τα προνόμια χορηγούνται ακριβώς όπως θα έπρεπε, με βάση μηχανισμούς εξουσιόδοτησης. Επιπλέον κάθε φορά που υπάρχει κάποια μεταβολή σε μια οντότητα, τότε τα προνόμιά της θα πρέπει να ανανεώνονται επειγόντως. Τέλος επιβάλλεται να αναγνωρίζονται άμεσα οι πιο επικίνδυνες οντότητες (δηλαδή αυτές με το μεγαλύτερο ρίσκο) και να ελαχιστοποιούνται τα προνόμιά τους αμέσως.
- **<u>Ταυτοποίηση πολλών παραγόντων</u>**: Οι υπεύθυνοι οργανισμοί για την εφαρμογή της εγκατάστασης της έξυπνης πόλης στην περιοχή τους, θα ήταν καλό να διασφαλίσουν τυχόν εφαρμογές εξ αποστάσεως σύνδεσης και να εφαρμόσουν την ταυτοποίηση πολλών παραγόντων σε τοπικές και απομακρυσμένες συσκευές και λογαριασμούς, όπου είναι εφικτό, έτσι ώστε να περιφρουρηθεί καλύτερα η υποδομή που επιτρέπει πρόσβαση στο οικοσύστημα της έξυπνης πόλης. Ειδικότερα, πρέπει να απαιτείται ταυτοποίηση πολλών παραγόντων σε περιοχές του

συστήματος όπου οι χρήστες έχουν κάνουν προνομιακά δικαιώματα ή έχουν πρόσβαση σε ευαίσθητες ή υψηλής αξίας πληροφορίες.

- **Αρχιτεκτονική Zero trust**: Εφαρμόζοντας πολιτικές ασφαλείας μηδενικής εμπιστοσύνης (zero trust) θα δημιουργηθεί ένα πιο ασφαλές οικοσύστημα έξυπνης πόλης, όπου θα χρειάζεται ταυτοποίηση και εξουσιοδότηση για κάθε νέα σύνδεση. Οι πολιτικές ασφαλείας zero trust επιτρέπουν μια μεγαλύτερη διαφάνεια και ορατότητα στις κινήσεις των δεδομένων και γενικότερα στη διακυβέρνηση του οικοσυστήματος της έξυπνης πόλης.
- **Διαχείριση αλλαγών**: Οι οργανισμοί που είναι υπεύθυνοι για την εφαρμογή της εγκατάστασης των τεχνολογιών της έξυπνης πόλης στην περιοχή τους, πρέπει να έχουν επίγνωση του συνολικού οικοσυστήματος και να διαχειρίζονται προσεκτικά τις επικοινωνίες μεταξύ των επιμέρους συστημάτων υποδομών, συμπεριλαμβανομένων των νέων συνδέσεων με υποδίκτυα που δημιουργούνται κατόπιν αλλαγών εσωτερικής αρχιτεκτονικής του οικοσυστήματος της έξυπνης πόλης. Οι διαχειριστές δικτύου θα πρέπει να παρακολουθούν διαρκώς τη δικτυακή υποδομή και να διασφαλίζουν την εύρυθμη και ασφαλή λειτουργία της.
- **Ενημερώσεις συστημάτων και εφαρμογών σε τακτικά διαστήματα**: Όπου είναι εφικτό να γίνεται αυτόματη ενημέρωση (σε επίπεδο λογισμικού ή/και υλισμικού) όλων των συσκευών που εμπλέκονται με τον έλεγχο πιστοποίησης ταυτότητας και ακεραιότητας. Επίσης η αναβάθμιση της επίγνωσης σε θέματα ασφαλείας (threat intelligence) για την αναγνώριση ενεργών απειλών και την εγγύηση ότι οι υποδομές και τα επιμέρους συστήματα διατηρούνται ασφαλή. Επίσης αυτή η διαδικασία πρέπει να προνοεί και για λογισμικά (και γενικότερα οντότητες του οικοσυστήματος) που πλησιάζει το τέλος της υποστήριξής τους (end of life support) μιας και μπορεί να σταματήσουν οι αναβαθμίσεις από τους κατασκευαστές τους και να θέσουν σε κίνδυνο το οικοσύστημα της έξυπνης πόλης.
- **Προμήθεια από έμπιστες πηγές**: Γενικότερα, οι κοινότητες που υλοποιούν οικοσυστήματα έξυπνων πόλεων στον χώρο τους, θα πρέπει να προμηθεύονται υλισμικό, λογισμικό, υπηρεσίες διαδικτύου των πραγμάτων, κλπ., μόνο από επίσημα κανάλια προμηθευτών κι έμπιστες πηγές. Όσον αφορά τις υπηρεσίες, οι οργανισμοί καλό θα ήταν να έχουν υπόψιν τους κινδύνους που μπορεί να διατρέχουν. Επιπλέον τα συμβόλαια παροχής υπηρεσιών (Service Level Agreements) που συνάπτονται με managed service providers ή cloud service providers, θα πρέπει να αναλύονται εξονυχιστικά, με κάθε λεπτομέρεια.
- **Αντίγραφα ασφαλείας συστημάτων και δεδομένων**: Οι αρμόδιοι οργανισμοί που εφαρμόζουν λύσεις σε οικοσυστήματα έξυπνων πόλεων θα πρέπει να διατηρούν αντίγραφα ασφαλείας, είτε τοπικά (σε απομονωμένα μέρος του οικοσυστήματος), είτε απομακρυσμένα σε κέντρα δεδομένων υπολογιστικών νεφών και να ελέγχουν περιοδικά, για την ορθή λειτουργία εφαρμογής σχεδίων ανάκτησης κι ανάκαμψης,

τόσο για τα αρχεία συστήματος, όσο και για τα σημαντικά δεδομένα τους. Αυτοί οι αρμόδιοι πρέπει να αποφασίσουν πως και που αυτά τα δεδομένα θα συλλέγονται, επεξεργάζονται, αποθηκεύονται και μεταδίδονται έτσι ώστε να μην υπάρχουν κενά στην ασφάλεια και να υπάρχει οργάνωση.

- **Εκπαίδευση προσωπικού**: Τα οικοσυστήματα των έξυπνων πόλεων περιλαμβάνουν ένα μεγάλο βαθμό αυτοματοποίησης, αλλά υπάρχει ανάγκη εκπαίδευσης του προσωπικού που τα διαχειρίζεται και τα λειτουργεί σε καθημερινή βάση στην ορθή εφαρμογή των πολιτικών ασφαλείας. Οι διαχειριστές αυτών των οικοσυστημάτων θα πρέπει να είναι σε θέση να απομονώνουν άμεσα επιμέρους συστήματα που έχουν εκτεθεί σε επιθέσεις και να εφαρμόζουν τα κατάλληλα σχέδια αντιμετώπισης κινδύνων κι ανάκαμψης από κακόβουλες επιθέσεις.
- **Αντιμετώπιση και αποκατάσταση:** Η ανάπτυξη και πρακτική των σχεδίων αντιμετώπισης συμβάντων και αποκατάστασης είναι πολύ κρίσιμη για μια ομαλή και αποτελεσματική αντιμετώπιση. Έτσι όπως γίνεται για παράδειγμα, με τις ασκήσεις σεισμού, πρέπει να καθιερωθεί η σχεδίαση και εφαρμογή κατάλληλων σχεδίων αντιμετώπισης κινδύνων κι ανάκαμψης από κακόβουλες επιθέσεις στο οικοσύστημα της έξυπνης πόλης.

3. Επιθέσεις Κοινωνικής Μηχανικής & Ηλεκτρονικό Ψάρεμα

3.1 Τι είναι Επιθέσεις Κοινωνικής Μηχανικής

Οι επιθέσεις κοινωνικής μηχανικής μπορούν να συμβούν σε μία στιγμή, αλλά μπορεί και να χρειαστούν μήνες για τη συλλογή επαρκών πληροφοριών σχετικά με τον στόχο προτού πραγματοποιηθεί επιτυχώς η πραγματική χειραγώγηση. Οι ίδιες τακτικές μπορούν να χρησιμοποιηθούν σε ένα και μόνο τηλεφώνημα ή σε ένα e-mail και η επίθεση να τελειώσει προτού το καταλάβετε. Οι επιθέσεις κοινωνικής μηχανικής συνήθως στοχεύουν στη χειραγώγηση των θυμάτων χρησιμοποιώντας τα εξής:

- **Συναισθήματα**: Χρησιμοποιούν τον φόβο ή τον ενθουσιασμό για να χειραγωγήσουν το θύμα ώστε να προβεί σε ενέργειες που συνήθως δεν θα έκανε. Όλα τα συναισθήματα μπορούν να χρησιμοποιηθούν ανάλογα με τους σκοπούς του επιτιθέμενου.
- **Αίσθηση ανάγκης**: Πείθουν το θύμα να σκεφτεί ότι πρέπει να ενεργήσει γρήγορα, με συνέπεια να αγνοήσει τις υποψίες σας.
- **Εμπιστοσύνη**: Οι περισσότεροι από αυτούς τους επιτιθέμενους είναι καλοί στο να κερδίζουν την εμπιστοσύνη των θυμάτων τους. Οι άνθρωποι εμπιστεύονται έμφυτα τους συναδέλφους τους και η επίθεση της κοινωνικής μηχανικής μπορεί να έρθει από κάποιον που νομίζουμε ότι γνωρίζουμε ή θα έπρεπε να γνωρίζουμε;

Αυτές οι δεξιότητες και τακτικές χρησιμοποιούνται συχνά με χιλιάδες διαφορετικούς τρόπους και για άλλα είδη επιθέσεων, δεν περιορίζονται μόνο στις επιθέσεις κοινωνικής μηχανικής.

Εικόνα 8. Πώς εκτίθεσαι σε επιθέσεις κοινωνικής μηχανικής;

Πηγή: https://static.safetydetectives.com/wp-content/uploads/2019/04/Greek.jpg

3.2 Τι είναι Ηλεκτρονικό Ψάρεμα

Το phishing (ηλεκτρονικό «ψάρεμα») βασίζεται στην εξαπάτηση του θύματος ώστε να εμπιστευτεί τον επιτιθέμενο και να του παράσχει τις απαραίτητες πληροφορίες. Η μορφή της επίθεσης μπορεί να είναι οτιδήποτε, όμως η πιο συχνή χρήση της επίθεσης phishing είναι μέσω e-mail. Τα e-mail, τα τηλεφωνήματα, τα SMS ή άλλες μορφές άμεσης επαφής μπορούν επίσης να χρησιμοποιηθούν για τη συλλογή περισσότερων πληροφοριών για επιθέσεις σε μεταγενέστερο χρόνο. Οποιεσδήποτε πληροφορίες παράσχει, το θύμα στον επιτιθέμενο, εκούσια ή ακούσια, θα τον βοηθήσουν να επιτύχει τις κακόβουλες δραστηριότητές του. Οι επιτιθέμενοι μπορούν να χρησιμοποιήσουν ακόμα και την παραμικρή πληροφορία που λαμβάνουν από το θύμα και να τη συνδυάσουν με άλλες πηγές πληροφοριών προκειμένου να επιτύχουν τους στόχους τους. Με τις πληροφορίες μπορούν, για παράδειγμα, να διαπράξουν κλοπή ταυτότητας ή απάτες, να λάβουν δάνεια ή να πείσουν τους συναδέλφους ή τους συγγενείς του θύματος να τους δώσουν πιο χρήσιμες πληροφορίες. Μια παραλλαγή του ηλεκτρονικού ψαρέματος είναι και το spear phishing (στοχευμένο ηλεκτρονικό «ψάρεμα»). Στην πραγματικότητα είναι μια εξειδικευμένη μορφή phishing που είναι πιο στοχευμένη. Πριν από τις επιθέσεις spear phishing συνήθως προηγείται η συλλογή πληροφοριών. Οι ενέργειες που αναφέρθηκαν προηγουμένως μπορούν να γίνουν με τις ίδιες μεθόδους phishing.

Οι πιο κοινές μέθοδοι που χρησιμοποιούνται στις επιθέσεις phishing είναι οι εξής:

- **Υπάρχει μια αίσθηση επείγοντος.** Μπορούν να χρησιμοποιηθούν διάφορες τακτικές που θα προσπαθήσουν να σας κάνουν να δράσετε γρήγορα, προτού σκεφτείτε. Για παράδειγμα, μια προσφορά περιορισμένου χρόνου ή κάτι ιδιαίτερο για τους πέντε πρώτους που θα παραγγείλουν. Μια άλλη επιλογή είναι η χρήση του φόβου. Η επίθεση μπορεί να προσπαθήσει να χρησιμοποιήσει τον φόβο ότι θα χάσετε τις προσωπικές σας οικονομικές πληροφορίες αν δεν ενεργήσετε γρήγορα. Η διατύπωση θα μπορούσε να είναι κάπως έτσι: «ο λογαριασμός σας βρίσκεται σε κίνδυνο, ενεργήστε τώρα για να τον προστατέψετε!»
- **Σύνδεσμοι.** Οι σύνδεσμοι σε αυτά τα e-mail μπορεί να μην είναι αυτό που δείχνουν. Χρησιμοποιούνται μέθοδοι που μπερδεύουν τις διευθύνσεις URL. Στα e-mail σε HTML, το κείμενο του συνδέσμου μπορεί να είναι οτιδήποτε και ο πραγματικός σύνδεσμος κάτι διαφορετικό. Οι χαρακτήρες μπορούν να αντικατασταθούν ώστε να μοιάζουν παρόμοιοι άλλα να κατευθύνουν σε διαφορετικές τοποθεσίες. Συνιστάται να πληκτρολογείτε τη διεύθυνση του δικτυακού τόπου μόνοι σας από το να κάνετε κλικ στον σύνδεσμο που λάβατε.
- **Συνημμένα αρχεία.** Τα συνημμένα αρχεία μπορούν να περιέχουν malware που θα επιτρέψει στους επιτιθέμενους να αποκτήσουν πρόσβαση στον υπολογιστή σας. Αυτό είναι ο πιο συχνός τρόπος που διασπείρεται το malware. Αν δεν περιμένετε

συνημμένα αρχεία, μην τα ανοίγετε. Διατηρείτε τους σαρωτές ιών ενημερωμένους με τους τελευταίους ορισμούς ιών.

- **Το μήνυμα είναι πολύ καλό για να είναι αληθινό.** Είτε έχετε κερδίσει κάτι ή έχετε καλές πιθανότητες να κερδίσετε. Η διατύπωση το κάνει να ακούγεται ως μια πολύ καλή προσφορά μίας ευκαιρίας, μόνο για εσάς!
- **Ο αποστολέας είναι ύποπτος.** Αν το e-mail είναι από γνωστό αποστολέα αλλά το περιεχόμενο είναι ασυνήθιστο, προσέξτε - το μήνυμα μπορεί να είναι phishing ή ανεπιθύμητη ηλεκτρονική αλληλογραφία. Επιβεβαιώστε ότι η διεύθυνση e-mail ταιριάζει με το όνομα του αποστολέα. Συχνά το όνομα μπορεί να είναι γνώριμο όμως η πραγματική διεύθυνση e-mail να είναι τυχαία. Αν το μήνυμα είναι από άγνωστο αποστολέα, να είστε ακόμα πιο καχύποπτοι.

Συνοπτικά , η επίθεση phishing είναι μια από τις πιο συχνές κακόβουλες επιθέσεις στο σύγχρονο ψηφιακό κόσμο, όπου η δραστηριότητα γίνεται ανώνυμα και έχει ως αποτέλεσμα την απώλεια προσωπικών δεδομένων, διαπιστευτηρίων και λοιπών ευαίσθητων πληροφοριών. Οι χρήστες δεν είναι σε θέση να αναγνωρίσουν τους τους κακόβουλους ή μη συνδέσμους και τους ιστότοπους από τους νόμιμους, επειδή μοιάζουν έχουν πλαστογραφηθεί με τέτοιο τρόπο ώστε να δείχνουν νομότυποι. Αυτό έχει σαν αποτέλεσμα, οι επιτιθέμενοι να προσπαθούν να επωφεληθούν από τα κενά ασφαλείας του κυβερνοχώρου. Οι έρευνες της Microsoft Consumer Safety Index Surveys αποκαλύπτουν ότι ο παγκόσμιος αντίκτυπος των επιθέσεων ηλεκτρονικού ψαρέματος με email phishing εκτιμάται σε περίπου 6 δισεκατομμύρια καθ’ έτος. Είναι δύσκολο να γνωρίζει κανείς πότε και πού οι επιτιθέμενοι επιχειρούν την επίθεση phishing. Με την πάροδο των ετών διάφορες τεχνικές βοήθησαν στον αποτελεσματικότερο εντοπισμό πολλών τέτοιων επιθέσεων phishing. Σήμερα διάφοροι αλγόριθμοι μηχανικής μάθησης είναι σε θέση να ανιχνεύουν, σε πραγματικό χρόνο, τις κακόβουλες επιθέσεις phishing με μεγαλύτερη ακρίβεια.

3.2.1 Διαφορετικοί τύποι επιθέσεων Phishing

Οι ακόλουθοι είναι οι πιο χαρακτηριστικοί τύποι επιθέσεων phishing:

- **Phishing βάσει αλγορίθμου**: Αυτός ο τύπος επίθεσης phishing δημιουργήθηκε με τη χρήση κατάλληλου αλγορίθμου και χρησιμοποιήθηκε για την αντιστοίχιση του αριθμού της πιστωτικής ή χρεωστικής κάρτας του λογαριασμού που εντοπίστηκε από την American Online (AOL).
- **Παραπλανητικό phishing**: Σε αυτόν τον τύπο επίθεσης phishing, οι χρήστες του διαδικτύου εξαπατώνται μέσω παραπλανητικών μηνυμάτων που τους καλούν να επαληθεύσουν τα στοιχεία των λογαριασμών τους και στη συνέχεια στέλνουν συνδέσμους που ζητούν την εισαγωγή (login/signup) των προσωπικών διαπιστευτηρίων τους προκειμένου να εγγραφούν ή να συνδεθούν στο

χειραγωγημένο σύνδεσμο. Με αυτό τον τρόπο αποσπούν τις ευαίσθητες πληροφορίες των θυμάτων.

- **URL Phishing:** Αυτή η ενέργεια γίνεται μέσω κρυφής διεύθυνσης URL. Το κρυφό URL περιέχει ιστοσελίδες phishers. Όταν ο χρήστης επισκεφθεί την πλαστογραφημένη διεύθυνση URL, οδηγείται στους διακομιστές του Phisher, όπου καταγράφονται κι αποθηκεύονται τα προσωπικά δεδομένα του θύματος.
- **Δηλητηρίαση Domain Name System server**: Περιλαμβάνει την εισαγωγή νέων καταχωρήσεων σε DNS servers για ψευδεπίγραφους ιστότοπους, έτσι ώστε όταν το θύμα επιχειρήσει να αποστείλει την αίτηση για επίσκεψη σε νομότυπο ιστότοπο, είτε να ανακατευθύνεται σε άλλον ψευδεπίγραφο ιστότοπο, είτε απλά να επιστρέφεται το σφάλμα "PAGE NOT FOUND".
- **Content Injection Phishing**: Οι χάκερς παρουσιάζουν τους ψεύτικους ιστότοπους ως νόμιμους. παραπλανούν τον χρήστη εξασφαλίζοντας ψευδείς ιστότοπους ως νόμιμους ιστότοπους (δηλ. γνωστό ως content spoofing). Αφού ο χρήστης ανακατευθυνθεί σε ψευδεπίγραφους ιστότοπους, μπορεί άθελά του να δώσει ευαίσθητες πληροφορίες του και ο επιτιθέμενος να τις αποθηκεύει στους διακομιστές του.
- **Whalephishing:** Αυτοί οι τύποι επιθέσεων phishing επικεντρώνονται σε ανώτερες αρχές ενός οργανισμού ή μιας εταιρείας για να διαρρεύσουν, ακούσια, τις εσωτερικές και ευαίσθητες πληροφορίες ενός οργανισμού.
- **Spear Phishing:** Απευθύνεται σε ένα συγκεκριμένο άτομο ή οργανισμό. Ο πρώτος επιτιθέμενος στέλνει το μήνυμα ηλεκτρονικού ταχυδρομείου και περιμένει τις απαντήσεις. Προσποιούνται ότι είναι αυθεντικό πρόσωπο ενός νόμιμου οργανισμού, ώστε να μπορέσουν να αποσπάσουν χρήσιμες και ευαίσθητες πληροφορίες.

3.2.2. Τρόποι ανίχνευσης επιθέσεων Phishing

Ακολουθούν οι διάφορες τεχνικές που έχουν προσαρμοστεί για τον εντοπισμό ύποπτων επιθέσεων ηλεκτρονικού ψαρέματος:

- **HTTP και HTTPS:** Το Https είναι πιο ασφαλές γιατί κρυπτογραφεί το μήνυμα από τον αποστολέα στον παραλήπτη και δεν υπάρχει διείσδυση ώστε οι επιτιθέμενοι να μπορούν να εκμεταλλευτούν τις πληροφορίες.
- **Πρόσβαση στο περιεχόμενο:** Το περιεχόμενο που υπάρχει στη διεύθυνση URL είναι προσβάσιμο στο χρήστη ή όχι.
- **Ανάγνωση των online κριτικών:** Αξιολόγηση των συγκεκριμένων URLs πώς μπορούμε να βεβαιωθούμε ότι η ιστοσελίδα στην οποία αποκτούμε πρόσβαση είναι νόμιμη και αξιόπιστη ή όχι.

- **Έλεγχος του ποιος είναι ο ιδιοκτήτης της ιστοσελίδας:** Πρέπει να ελέγξουμε και να μάθουμε το ιστορικό της συγκεκριμένης ιστοσελίδας (π.χ. τον ιδιοκτήτη, τη βαθμολογία της) πριν αποκτήσουμε πρόσβαση.
- **Προβολή λεπτομερειών πιστοποίησης:** Βεβαιωθείτε ότι ο συγκεκριμένος ιστότοπος είναι νόμιμα πιστοποιημένος από τον αρμόδιο οργανισμό του ή όχι.

3.2.3. Επιθέσεις Ηλεκτρονικού Ψαρέματος σε Websites

Η επίθεση phishing είναι ένα είδος ολοκληρωμένης εξαπάτησης – δραστηριότητας που συμβαίνει όταν οι ψεύτικοι ιστότοποι συμπεριφέρονται ως νόμιμοι προκειμένου να εξαγάγουν και να αποκτήσουν τα δεδομένα και τις πληροφορίες, όπως για παράδειγμα τον κωδικό πρόσβασης, το e-mail και το σημείο του λογαριασμού, όπως ο λογαριασμός διαπιστευτηρίων.

Ειδικότερα , η επίθεση μπορεί να διεξάγει το έργο της με ανώνυμο τρόπο. Η επίθεση phishing ως επί το πλείστον ο αθώος χρήστης χάνει τα ευαίσθητα, μοναδικά, προσωπικά, πολύτιμα και ασφαλή δεδομένα και πληροφορίες. Πολλοί χάκερς δουλεύουν μέσω επιθέσεων phishing όπου οι πελάτες παγιδεύονται σε αλληλεπίδραση με ιστοσελίδες που μοιάζουν να είναι νόμιμες ιστοσελίδες .Το phishing είναι μια σημαντική απειλή για όλους τους ψηφιακούς χρήστες στον κόσμο του κυβερνοχώρου . Κάθε εργασία γίνεται στο διαδίκτυο και υπάρχει μεγάλη πιθανότητα να αξιοποιηθούν προσωπικές πληροφορίες και διαπιστευτήρια. Οι Phishers άρχισαν να κερδίζουν μετρητά και το κάνουν αυτό ως μια καρποφόρα επιχείρηση, χρησιμοποιούν διάφορες τεχνικές για να επιτεθούν στους αδύναμους πελάτες, όπως ενημέρωση, VOIP, ψεύτικη σύνδεση και ψεύτικες τοποθεσίες. Δεν είναι δύσκολο να δημιουργηθούν ψεύτικοι ιστότοποι , οι οποίοι μοιάζουν με έναν μέσο ιστότοπο όσον αφορά τη μορφή και το περιεχόμενο.

Για να διακρίνει κανείς μια ιστοσελίδα phishing υπάρχουν διάφοροι τρόποι για να την βρει, όπως το HTTPS είναι πιο ασφαλές από το HTTP επειδή εξασφαλίζει την τριάδα CIA μεταξύ αποστολέα και παραλήπτη. Το περιεχόμενο του ιστότοπου, η ηλεκτρονική επανεξέταση, ο έλεγχος του ιστορικού του ιστότοπου και τα στοιχεία πιστοποίησης είναι οι προηγούμενοι τρόποι διάκρισης του phishing και του νόμιμου ιστότοπου. Σημαντικό μειονέκτημα αυτής της τεχνικής είναι ότι, δεν μπορεί να αναγνωρίσει την επίθεση phishing party time. Καθώς τα περισσότερα μηνύματα phishing είναι διατεταγμένα να εμφανίζονται από μια γνήσια πηγή, ένα τεράστιο επίπεδο σάρωσης του πελάτη ηλεκτρονικού ταχυδρομείου αντιλαμβάνεται την επίθεση phishing. Η εκτέλεση του Secure Socket Layer (SSL) και η προηγμένη πιστοποίηση ταυτότητας (CA) δεν διασφαλίζει επιπλέον τον πελάτη του διαδικτύου έναντι αυτής της επίθεσης. Στην επίθεση παρωδίας ιστού, ο επιτηρητής κατευθύνει την πρόσκληση σε παραποιημένο εργαζόμενο στον ιστό. Η αλήθεια είναι ότι μπορεί να κατασκευαστεί ένα συγκεκριμένο είδος SSL και CA, ενώ όλα αυτά δίνουν την εντύπωση ότι είναι αυθεντικά. Όπως υποδεικνύεται, η ασφαλής ένωση ανάγνωσης δεν

κάνει πρακτικά τίποτα για να προστατεύσει τους πελάτες ιδιαίτερα από τους επιτήδειους που έχουν πληροφορίες σχετικά με το πώς λειτουργούν οι "ασφαλείς" ενώσεις.

Η διεύθυνση URL είναι μια παγκόσμια τοποθεσία του κορδονιού στον Παγκόσμιο Ιστό, και συμπληρώνει τον βασικό τρόπο εύρεσης μιας έκθεσης στον Ιστό. Πράγματι, ακόμη και σε περιπτώσεις όπου η ουσία των ιστότοπων αντιγράφεται, η διεύθυνση URL θα μπορούσε σε κάθε περίπτωση να χρησιμοποιηθεί για την αναγνώριση γνήσιων ιστότοπων από τις απάτες. Στην επεξεργασία δεδομένων έχουμε λάβει το σύνολο δεδομένων δομής. Το φίλτρο και το περιτύλιγμα είναι η τεχνική που βοηθά στη διόρθωση του συνόλου δεδομένων, διορθώνει επίσης τις τιμές δεδομένων που λείπουν και επίσης, εξαλείφει το σύνολο δεδομένων που δεν έχουν καμία σχέση με την ακρίβεια. Βοηθά στον καθορισμό, τον εντοπισμό και τη διόρθωση σφαλμάτων στο δεδομένο σύνολο δεδομένων και βοηθά επίσης στην εξαγωγή σημαντικού και ειδικού περιεχομένου που είναι σχετικό με την ανάλυση και τον υπολογισμό. Μετά την προ-επεξεργασία των δεδομένων εξάγουμε συγκεκριμένα χαρακτηριστικά εισόδου από το σύνολο δεδομένων εισόδου. Με τον τρόπο αυτό μειώνουμε τη διάσταση του συνόλου δεδομένων και αφαιρούμε τον πλεονασμό των χαρακτηριστικών. Μερικά από τα χαρακτηριστικά που εξάγονται είναι χαρακτηριστικά γραμμής διευθύνσεων, ανώμαλα βασικά χαρακτηριστικά, βασικά χαρακτηριστικά html και java script και βασικά χαρακτηριστικά τομέα.

3.3 Αντίμετρα σε Επιθέσεις Ηλεκτρονικού Ψαρέματος

Λαμβάνοντας διάφορα τηλεφωνήματα και μηνύματα απάτης και αντιμετωπίζοντας διάφορες ιστοσελίδες για απάτες. Στόχος είναι να ελαχιστοποιηθεί η επίδραση του προβλήματος. Το πρόβλημα που προκαλείται από την ιστοσελίδες phishing έχει ως αποτέλεσμα την απώλεια προσωπικών δεδομένων, οικονομική απώλεια ή εκβιασμό. Αυτοί που είναι εξοικειωμένοι με τις phishing ιστοσελίδες μπορούν να αναγνωρίσουν τα σημάδια και να τα ξεπεράσουν, αλλά εκείνοι με λιγότερες γνώσεις μπορεί να είναι τα θύματα. Έτσι έχουμε σχεδιάσει να δημιουργήσουμε User Interface για την ανίχνευση τέτοιων sites. Η ανίχνευση και η πρόληψη phishing ιστοσελίδων είναι όλο και πιο σημαντική από το παρελθόν στο μέλλον μέρα με τη μέρα. Διάφορα είδη των στρατηγικών phishing προσφέρουν θεμελιώδεις τρόπους αντιμετώπισης για την αστυνομία, τις εταιρίες αλλά και το άτομο.

- **Εκπαίδευση εργαζομένων:** Η εκπαίδευση των εργαζομένων αποτελεί βασικό πυλώνα για την ενίσχυση της άμυνας ενός οργανισμού έναντι των επιθέσεων phishing. Η επένδυση σε ολοκληρωμένα προγράμματα κατάρτισης δίνει τη δυνατότητα στους υπαλλήλους να αναγνωρίζουν και να ανταποκρίνονται αποτελεσματικά σε απόπειρες phishing. Οι διαδραστικές συνεδρίες, οι ασκήσεις προσομοίωσης phishing και οι συνεχείς εκστρατείες ευαισθητοποίησης συμβάλλουν στην καλλιέργεια ενός επαγρυπνούντος και ευαισθητοποιημένου σε θέματα ασφάλειας εργατικού δυναμικού.

- **Φιλτράρισμα ηλεκτρονικού ταχυδρομείου:** Η εφαρμογή ισχυρών μηχανισμών φιλτραρίσματος ηλεκτρονικού ταχυδρομείου χρησιμεύει ως κρίσιμη γραμμή άμυνας έναντι των επιθέσεων phishing. Οι προηγμένες λύσεις φιλτραρίσματος ηλεκτρονικού ταχυδρομείου χρησιμοποιούν αλγορίθμους μηχανικής μάθησης και αναγνώριση προτύπων για τον εντοπισμό και την απομόνωση ύποπτων μηνυμάτων ηλεκτρονικού ταχυδρομείου. Αποκλείοντας τα κακόβουλα συνημμένα αρχεία και τις διευθύνσεις URL, οι οργανισμοί μπορούν να μειώσουν σημαντικά την πιθανότητα επιτυχημένων περιστατικών phishing μέσω των καναλιών ηλεκτρονικού ταχυδρομείου.
- **Μαύρες λίστες:** Η διατήρηση ενημερωμένων μαύρων λιστών γνωστών κακόβουλων τομέων και διευθύνσεων IP είναι επιτακτική ανάγκη για την προληπτική αποτροπή της πρόσβασης σε ιστότοπους phishing. Η ενσωμάτωση με τροφοδοσίες πληροφοριών απειλών διασφαλίζει ότι οι οργανισμοί μπορούν να εντοπίζουν και να αποκλείουν γρήγορα κακόβουλες οντότητες που επιχειρούν να διεισδύσουν στο δίκτυό τους. Η τακτική ενημέρωση των μαύρων λιστών ενισχύει την ανταπόκριση της υποδομής ασφαλείας στις εξελισσόμενες απειλές phishing.
- **Φιλτράρισμα DNS:** Το φιλτράρισμα DNS παίζει καθοριστικό ρόλο στην αποτροπή της πρόσβασης των χρηστών σε κακόβουλους ιστότοπους που χρησιμοποιούνται για phishing. Αναλύοντας τα αιτήματα DNS και αποκλείοντας τα αιτήματα προς γνωστούς τομείς phishing, οι οργανισμοί μπορούν να προσθέσουν ένα πρόσθετο επίπεδο προστασίας. Οι έξυπνες λύσεις φιλτραρίσματος DNS μπορούν επίσης να εντοπίζουν και να αποκλείουν αλγόριθμους δημιουργίας τομέων (DGA), μια τεχνική που χρησιμοποιείται συχνά από εκστρατείες phishing για να αποφεύγεται η ανίχνευση.
- **Ανάλυση της συμπεριφοράς των χρηστών:** Η ανάλυση της συμπεριφοράς των χρηστών περιλαμβάνει την παρακολούθηση και την ανάλυση της ψηφιακής συμπεριφοράς των εργαζομένων για τον εντοπισμό ανωμαλιών που μπορεί να υποδεικνύουν ένα πιθανό περιστατικό phishing. Με τον καθορισμό της βασικής συμπεριφοράς των χρηστών και τη χρήση προηγμένης ανάλυσης, οι οργανισμοί μπορούν να εντοπίζουν αποκλίσεις από τα συνήθη πρότυπα, επιτρέποντας την έγκαιρη παρέμβαση. Η συνεχής παρακολούθηση και ανάλυση συμβάλλουν σε μια προληπτική προσέγγιση για τον εντοπισμό και τον μετριασμό των απειλών phishing.
- **Αντιμετώπιση περιστατικών:** Ένα ισχυρό σχέδιο αντιμετώπισης περιστατικών είναι απαραίτητο για την ελαχιστοποίηση των επιπτώσεων των επιτυχημένων επιθέσεων phishing. Οι οργανισμοί θα πρέπει να αναπτύσσουν και να δοκιμάζουν τακτικά τις διαδικασίες αντιμετώπισης περιστατικών για να διασφαλίζουν ταχεία και συντονισμένη αντίδραση. Αυτό περιλαμβάνει την απομόνωση των επηρεαζόμενων συστημάτων, τη διεξαγωγή εγκληματολογικής ανάλυσης και την εφαρμογή διορθωτικών μέτρων για την αποτροπή μελλοντικών περιστατικών. Μια καλά προετοιμασμένη ομάδα αντιμετώπισης περιστατικών είναι ζωτικής σημασίας για τον αποτελεσματικό μετριασμό των συνεπειών μιας επίθεσης phishing.

4. Μηχανική Μάθηση

4.1 Εισαγωγή στη Μηχανική Μάθηση

Ο όρος Μηχανική Μάθηση αναφέρεται στον αυτόματο εντοπισμό ουσιώδης μοτίβων σε δεδομένα. Τις τελευταίες δεκαετίες, έχει γίνει ένα συνηθισμένο εργαλείο σχεδόν σε κάθε εργασία που χρειάζεται εξόρυξη πληροφοριών από μεγάλες συλλογές δεδομένων. Είμαστε περικυκλωμένοι από τεχνολογία που βασίζεται στην μηχανική μάθηση, ας δώσουμε μερικά παραδείγματα.

Μηχανές αναζήτησης μαθαίνουν πως να μας φέρνουν τα καλύτερα αποτελέσματα, λογισμικά anti-spam μαθαίνουν πως να φιλτράρουν τα μηνύματα ηλεκτρονικού ταχυδρομείου μας, και πιστωτικές συναλλαγές ασφαλίζονται από λογισμικό που μαθαίνει να ανακαλύπτει απόπειρες κοροϊδίας με σκοπό την διάσπαση χρήσιμων πληροφοριών, κάμερες ψηφιακές μαθαίνουν να αναγνωρίζουν πρόσωπα και ευφυής εφαρμογές στα τηλέφωνα μαθαίνουν να αναγνωρίζουν ανθρώπινες φωνές, αυτοκίνητα μπορούν να οδηγούν μόνα τους με την χρήση της μηχανικής μάθησης. Επίσης η μηχανική μάθηση είναι ευρέος γνωστή στις επιστημονικές εφαρμογές όπως για παράδειγμα αστρονομία, βιο-ιατρική και άλλα.

Μια συνηθισμένη λειτουργία όλων αυτών των εφαρμογών που χρησιμοποιούν μηχανική μάθηση, σε σχέση με την κλασσική χρήση του υπολογιστή , είναι η ανικανότητα στο να μάθουν ακριβώς το τι ζητάμε απευθείας, χωρίς καν να γίνει κάποια εκπαίδευση, αυτό οφείλεται κυρίως στην πολυπλοκότητα των μοτίβων. Όπως και εμείς οι άνθρωποι που είμαστε ευλογημένοι με ευφυΐα και σκέψη, έτσι η μηχανική μάθηση μας έχει ως παράδειγμα. Οποιαδήποτε τέχνη, γνώση ή δεξιότητα έχουμε, την έχουμε μάθει από εμπειρία και όχι επειδή κάποιος μας είπε πως είναι μια συγκεκριμένη τέχνη, γνώση, δεξιότητα, με τον ίδιο τρόπο η μηχανική μάθηση μαθαίνει από την εμπειρία που εμείς της δίνουμε έτσι ώστε να κάνει έπειτα την ανάλογη λειτουργία για την οποία έχει δημιουργηθεί. Ο σκοπός αυτής της ενότητας είναι να εξηγηθούν βασικές ορολογίες και μεθοδολογίες που θα βοηθήσουν παρακάτω.

4.1.1 Βασικές γνώσεις

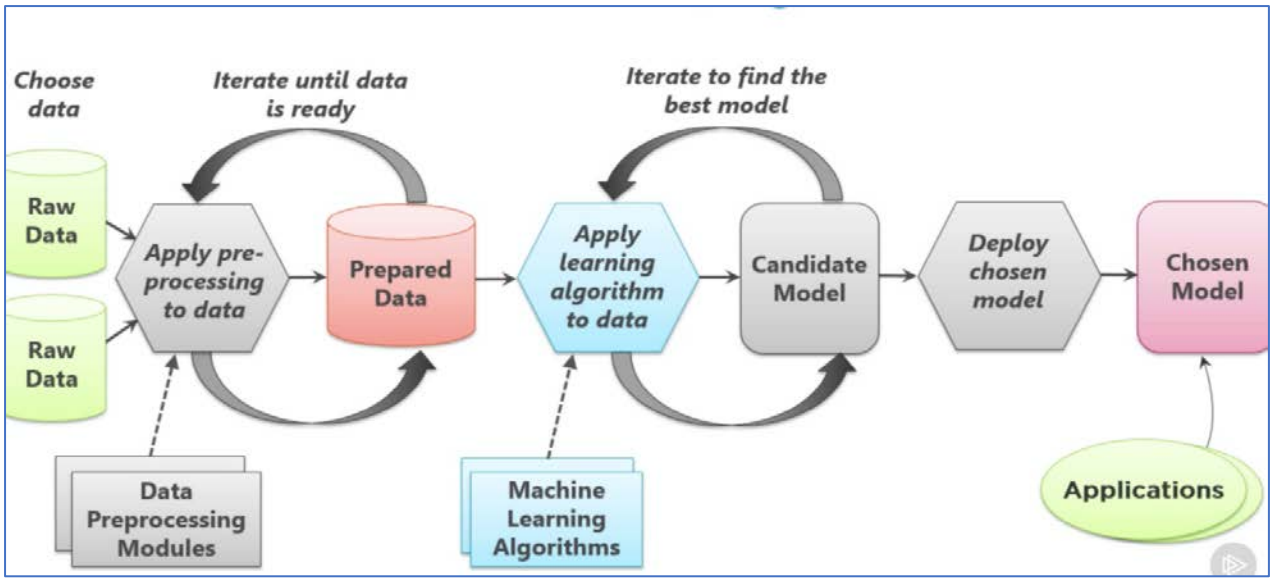
Ορισμός: Μηχανική Μάθηση [17] λέγεται η επιστήμη της πληροφορικής που εστιάζει στην χρήση δεδομένων και αλγόριθμων για να αντιγράψουν τον τρόπο με τον οποίο οι άνθρωποι μαθαίνουν βελτιώνοντας σταδιακά την ακρίβεια.

Βήματα αλγορίθμου: Αρχικά, η διαδικασία ξεκινάει με τη ταυτοποίηση του προβλήματος, με σκοπό την καλύτερη κατανόησή του. Στην συνέχεια ξεκινάει η συλλογή των δεδομένων που θα χρησιμοποιήσει το μοντέλο μηχανικής μάθησης για να εκπαιδευτεί, αυτά τα δεδομένα πρέπει να είναι κυρίως ορθά και από έμπιστες πηγές έτσι ώστε να μην υπάρχει παραπληροφόρηση στο μοντέλο με συνέπεια λάθος προβλέψεις. Αφού, έχουν μαζευτεί

όλες οι δυνατές πληροφορίες, μέρος παίρνει η προ επεξεργασία των δεδομένων. Αυτό γίνεται για να μπορέσει το μοντέλο να καταλάβει τι διαβάζει, παίρνοντας μια μικρή καθοδήγηση, και γενικότερα να διευκολυνθεί στην μάθηση. Όταν αυτό το βήμα ολοκληρωθεί, σειρά έχει η ανάπτυξη του αλγορίθμου στα μέτρα του προβλήματος που είναι προορισμένο να λύσει και τα δεδομένα που είναι έτοιμο να δεχτεί.

Δυστυχώς, ενώ θα μπορούσαμε απλά να είχαμε τελειώσει εδώ τα βήματα, θα πρέπει το αποτέλεσμα να αξιολογηθεί χρησιμοποιώντας κάποια μετρικά τα οποία θα αναφερθούν συγκεκριμένα παρακάτω. Ο λόγος που γίνεται η αξιολόγηση είναι για να γίνουν οι αλλαγές που χρειάζονται στο μοντέλο ή στην προ επεξεργασία των δεδομένων μιας και είναι απίθανο να δοθεί το επιθυμητό αποτέλεσμα με την πρώτη, εδώ φαίνεται και η ομοιότητα με την ανθρώπινη μάθηση για παράδειγμα όταν τα παιδιά κάνουν λάθη στο σχολείο και η δασκάλα τους εξηγεί τι να αλλάξουν για να είναι σωστοί και αυτό μας οδηγεί στο επόμενο βήμα. Συλλογή αξιολόγησης, σε αυτό το στάδιο συλλέγεται όλα τα χρήσιμα αποτελέσματα από την αξιολόγηση έτσι ώστε να υπάρχει μια γενικότερη εικόνα με το τι πάει λάθος [18].

Τέλος μετά από πολλές προσπάθειες να επιλεχθεί ο καλύτερος αλγόριθμος και θα δημιουργηθεί το μοντέλο με όλα τα αναγκαία πρακτικά πλαίσια και ως αποτέλεσμα το τελικό τεστ με τις προβλέψεις που παράχθηκαν.

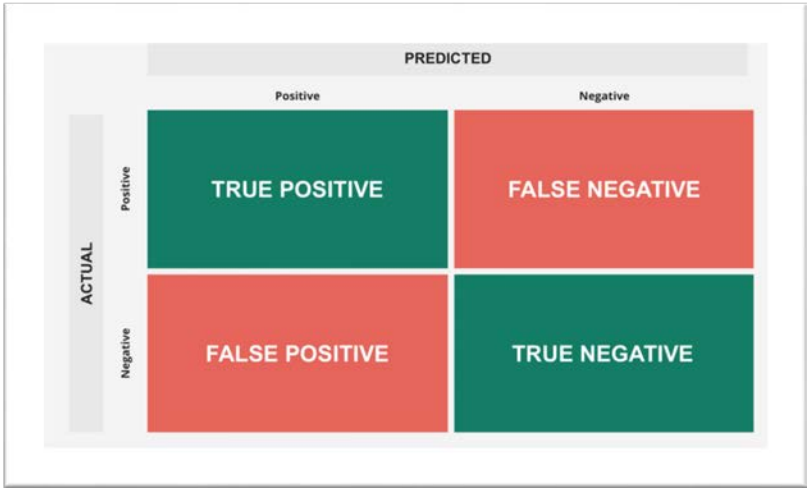


Εικόνα 9. Μεθοδολογία Μηχανικής Μάθησης

Πηγή: https://volcanohong.github.io/2016/09/01/machine-learning-notes/

Μετρήσεις: Υπάρχουν διαφορετικοί δείκτες και μέτρα για την αξιολόγηση των μοντέλων μηχανικής μάθησης. Σε γενικές γραμμές τα Confusion matrix είναι από τις πιο σημαντικές μετρήσεις και τρόπους για να δείξεις τι ακριβώς έκανε το μοντέλο. Περιέχει 4 μετρήσεις, πραγματικά θετικά, ψευδά θετικά, πραγματικά αρνητικά και ψευδά αρνητικά και ο συνδυασμός αυτών δημιουργεί πολλές άλλες μετρήσεις όπως θα δούμε παρακάτω. Αρχικά ας εξηγήσουμε τις τέσσερις αυτές βασικές μετρήσεις [19]:

- **True Positive:** Ο αριθμός θετικών δειγμάτων που το μοντέλο ταξινομεί σωστά.
- **False Positive:** Ο αριθμός αρνητικών δειγμάτων που το μοντέλο ταξινόμησε λάθος (Λάθος συναγερμός).
- **True Negative:** Ο αριθμός αρνητικών δειγμάτων που το μοντέλο ταξινόμησε σωστά.
- **False Negative:** Ο αριθμός θετικών δειγμάτων που το μοντέλο ταξινόμησε λάθος.



Εικόνα 10. Βασικές Μετρήσεις Αξιολόγησης Μοντέλων Μηχανικής Μάθησης

Πηγή: https://dataaspirant.com/wp-content/uploads/2020/08/3_confusion_matrix.png

Πίνακας 1. Βασικές Μετρήσεις Αξιολόγησης Μοντέλων Μηχανικής Μάθησης

TP	*TRUE POSITIVE*	*ΑΛΗΘΕΣ ΘΕΤΙΚΟ*
FP	*FALSE POSITIVE*	*ΨΕΥΔΕΣ ΘΕΤΙΚΟ*
TN	*TRUE NEGATIVE*	*ΑΛΗΘΕΣ ΑΡΝΗΤΙΚΟ*
FN	*FALSE NEGATIVE*	*ΨΕΥΔΕΣ ΑΡΝΗΤΙΚΟ*
PDV	*PRECISION/POSITIVE PREDICTIVE VALUE*	*ΑΚΡΙΒΕΙΑ/ ΘΕΤΙΚΗ ΠΡΟΓΝΩΣΤΙΚΗ ΑΞΙΑ*
TPR	*RECALL/SENSITIVITY/TRUE POSITIVE RATE*	*ΑΝΑΚΛΗΣΗ/ΕΥΑΙΣΘΗΣΙΑ/ ΑΛΗΘΩΣ ΘΕΤΙΚΟ ΠΟΣΟΣΤΟ*
TNR	*SPECIFICITY/TRUE NEGATIVE RATE*	*ΕΙΔΙΚΟΤΗΤΑ / ΑΛΗΘΩΣ ΑΡΝΗΤΙΚΟ ΠΟΣΟΣΤΟ*
FPR	*FALL OUT/FALSE POSITIVE RATE*	*ΠΤΩΣΗ / ΨΕΥΔΟΣ ΘΕΤΙΚΟ ΠΟΣΟΣΤΟ*
FNR	*MISS RATE/FALSE NEGATIVE RATE*	*ΠΟΣΟΣΤΟ ΑΣΤΟΧΙΑΣ / ΨΕΥΔΟΣ ΑΡΝΗΤΙΚΟ ΠΟΣΟΣΤΟ*
FDR	*FALSE DISCOVERY RATE*	*ΠΟΣΟΣΤΟ ΨΕΥΔΟΥΣ ΑΝΑΚΑΛΥΨΗΣ*
FOR	*FALSE OMISSION RATE*	*ΠΟΣΟΣΤΟ ΨΕΥΔΟΥΣ ΠΑΡΑΛΕΙΨΗΣ*
ROC	*RECEIVED OPERATING CHARACTERISTIC CURVE*	*ΛΑΜΒΑΝΟΜΕΝΗ ΛΕΙΤΟΥΡΓΙΚΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΗ ΚΑΜΠΥΛΗ*
AUC	*AREA UNDER CURVE*	*ΠΕΡΙΟΧΗ ΥΠΟ ΚΑΜΠΥΛΗ*
MSE	*MEAN SQUARED ERROR*	*ΜΕΣΟ ΤΕΤΡΑΓΩΝΙΚΟ ΣΦΑΛΜΑ*
MAE	*MEAN ABSOLUTE ERROR*	*ΜΕΣΟ ΑΠΟΛΥΤΟ ΣΦΑΛΜΑ*
MAPE	*MEAN ABSOLUTE PREDICTION ERROR*	*ΜΕΣΟ ΑΠΟΛΥΤΟ ΠΡΟΒΛΕΠΟΜΕΝΟ ΣΦΑΛΜΑ*

Και τώρα ας αναλύσουμε τους συνδυασμούς [19] [20] [21] [22] [23] [24] [25] [26] [27]:

a) **Precision:** Είναι ο λόγος των σωστά ταξινομημένων θετικών δειγμάτων προς όλα τα θετικά δείγματα της πρόβλεψης. Δείχνει το πόσα θετικά δείγματα προβλέπει το μοντέλο. Όσο μεγαλύτερη η τιμή τόσο καλύτερη απόδοση υπάρχει στο μοντέλο. Όταν το Precision κοντεύει προς το 1 σημαίνει πως το μοντέλο δεν έχασε πραγματικά θετικά (True Positives) και είναι ικανό να ταξινομεί σωστά ανάμεσα σε σωστά και λάθος δείγματα.

$$Precision = \frac{TP}{TP + FP}$$

b) **Recall:** Η αλλιώς και Sensitivity είναι το ποσοστό των θετικών δειγμάτων που έχουν ταξινομηθεί σωστά ως προς το σύνολο όλων των πραγματικών θετικών δειγμάτων. Όταν το recall κοντεύει προς το 1 σημαίνει πως το μοντέλο δεν έχασε πραγματικά θετικά (True Positives) και είναι ικανό να ταξινομεί σωστά ανάμεσα σε σωστά και λάθος θετικά δείγματα.

$$Recall = \frac{TP}{TP + FN}$$

c) **Specificity:** Είναι ο λόγος των σωστά ταξινομημένων αρνητικών δειγμάτων προς το σύνολο όλων των αρνητικών δειγμάτων στα δεδομένα. Δείχνει το πόσα αρνητικά δείγματα προβλέπει το μοντέλο. Όσο μεγαλύτερη η τιμή τόσο καλύτερη απόδοση υπάρχει στο μοντέλο.

$$Specificity = \frac{TN}{TN + FP}$$

d) **Accuracy:** Είναι ο λόγος των σωστά ταξινομημένων δειγμάτων προς όλα τα δείγματα του συνόλου δεδομένων. Δείχνει το πόσα σωστά δείγματα προβλέπει το μοντέλο. Όσο μεγαλύτερη η τιμή τόσο καλύτερη απόδοση υπάρχει στο μοντέλο.

$$Accuracy = \frac{TP + TN}{TN + FP + FN + TP}$$

e) **Error Rate:** Είναι ο λόγος των λάθος ταξινομημένων δειγμάτων προς όλα τα δείγματα του συνόλου δεδομένων. Δείχνει το πόσα λάθος δείγματα προβλέπει το μοντέλο. Όσο μικρότερη η τιμή τόσο καλύτερη απόδοση υπάρχει στο μοντέλο.

$$Error\ Rate = \frac{FP + FN}{TN + FP + FN + TP}$$

f) **Fall Out:** Είναι ο λόγος των λάθος ταξινομημένων αρνητικών δειγμάτων προς όλα τα αρνητικά δείγματα του συνόλου δεδομένων. Δείχνει το πόσα αρνητικά δείγματα προβλέπει το μοντέλο ως θετικά. Όσο μικρότερη η τιμή τόσο καλύτερη απόδοση υπάρχει στο μοντέλο.

$$Fall\ Out = \frac{FP}{FP + TN}$$

g) **Miss Rate:** Είναι ο λόγος των λάθος ταξινομημένων θετικών δειγμάτων προς όλα τα θετικά δείγματα του συνόλου δεδομένων. Δείχνει την το πόσα θετικά δείγματα προβλέπει το μοντέλο ως αρνητικά. Όσο μικρότερη η τιμή τόσο καλύτερη απόδοση υπάρχει στο μοντέλο.

$$Miss\ Rate = \frac{FN}{FN + TP}$$

h) **FDR:** Είναι ο λόγος των λάθος ταξινομημένων αρνητικών δειγμάτων προς όλα τα ταξινομημένα αρνητικά δείγματα του συνόλου δεδομένων. Δείχνει το πόσα θετικά δείγματα προβλέπει το μοντέλο ως αρνητικά. Όσο μικρότερη η τιμή τόσο καλύτερη απόδοση υπάρχει στο μοντέλο.

$$False\ Discovery\ Rate = \frac{FP}{FP + TP}$$

i) **FOR:** Είναι ο λόγος των λάθος ταξινομημένων θετικών δειγμάτων προς όλα τα ταξινομημένα θετικά δείγματα του συνόλου δεδομένων. Δείχνει την πιθανότητα η πραγματική τιμή να είναι θετική. Όσο μικρότερη η τιμή τόσο καλύτερη απόδοση υπάρχει στο μοντέλο.

$$False\ Omission\ Rate = \frac{FN}{FN + TN}$$

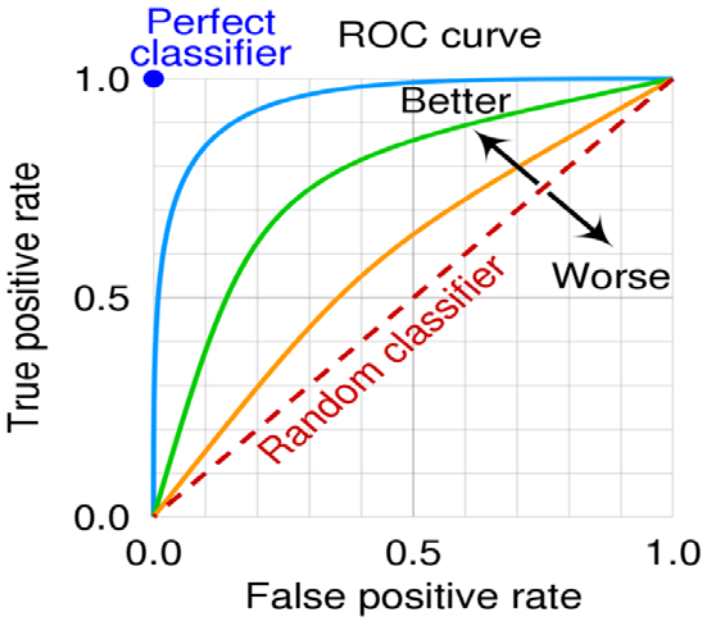
j) **F1-Score:** Είναι ένας συνδυασμός των Precision και Recall, γενικότερα είναι ένας αρμονικός συνδυασμός των δύο. Ένα υψηλό F1 συμβολίζει ένα υψηλό Precision και ένα υψηλό Recall, δίνοντας μια καλή ισορροπία σε προβλήματα ανώμαλης ταξινόμησης.

$$F1\ Score = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

k) **G-Mean:** Υπολογίζεται χρησιμοποιώντας τις πραγματικές προβλέψεις που ταξινομούνται. Στην περίπτωση που τα αρνητικά δείγματα είναι περισσότερα από τα θετικά δείγματα το accuracy δεν θα μας βοηθήσει στην γενική εικόνα του μοντέλου, έτσι χρησιμοποιούμε το G-Mean.

$$G\ MEAN = \sqrt{\frac{TP}{TP + FN} \times \frac{TN}{TN + FP}}$$

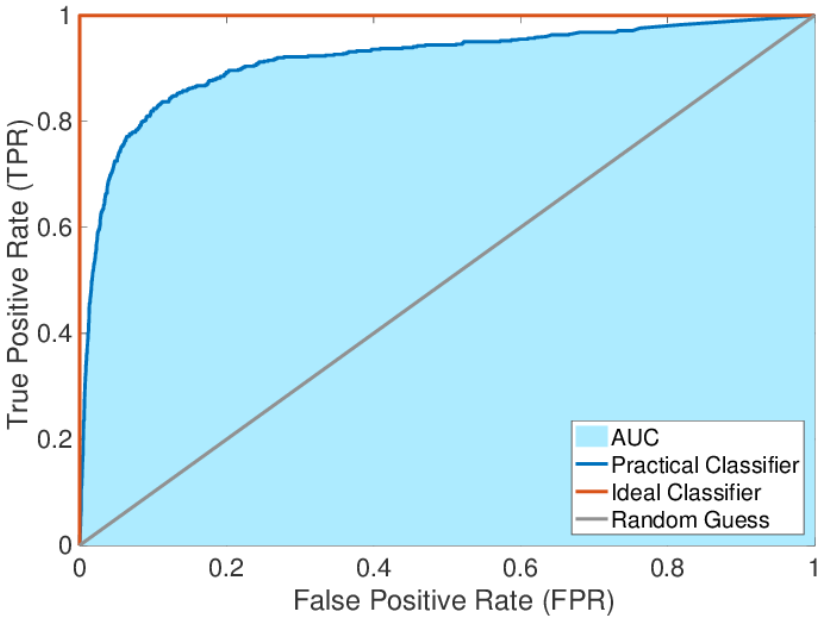
l) **ROC:** Είναι ένα γράφημα που δείχνει την απόδοση ενός μοντέλου ταξινόμησης στο σύνολο των ταξινομήσεων, βάζει στον έναν άξονα το Recall και στον άλλον άξονα το Fall out. Όσο πιο κοντά το γράφημα στην πάνω αριστερή γωνία τόσο καλύτερη απόδοση υπάρχει στο μοντέλο.



Εικόνα 11. Receiver Operating Characteristic (ROC)

Πηγή: https://www.nomidl.com/machine-learning/what-is-the-roc-curve/

m) **AUC:** Είναι ένα γράφημα που δείχνει την απόδοση ενός μοντέλου ταξινόμησης στο σύνολο των ταξινομήσεων, βάζει τα True Positive στον έναν άξονα και τα False Positive στον άλλον. Υπολογίζει όλη την επιφάνεια κάτω από το ROC. Είναι η πιθανότητα το μοντέλο να ταξινόμηση ένα τυχαίο δείγμα θετικό παρά αρνητικό Όσο μεγαλύτερη η τιμή τόσο καλύτερη απόδοση υπάρχει, ένα μοντέλο με 100% σωστές προβλέψεις θα έχει τιμή AUC 1 [24].



Εικόνα 12. Area Under Curve (AUC)

Πηγή:https://www.researchgate.net/figure/Receiver-Operating-Characteristic-ROC-curves-and-the-area-under-ROC-curve-or-AUC_fig3_331797273

n) **MSE:** Είναι ένας γραμμικός υπολογισμός του σφάλματος όπου η τιμή της διαφοράς μεταξύ των προβλεπόμενων τιμών και των πραγματικών τιμών είναι τετραγωνισμένη. Είναι ο τετραγωνισμένος συνολικός μέσος όρος όλου του μοντέλου. Όσο μικρότερη η τιμή τόσο καλύτερη απόδοση υπάρχει.

$$\text{Mean Square Error} = \frac{\sum (\hat{y}_i - y_i)^2}{n}$$

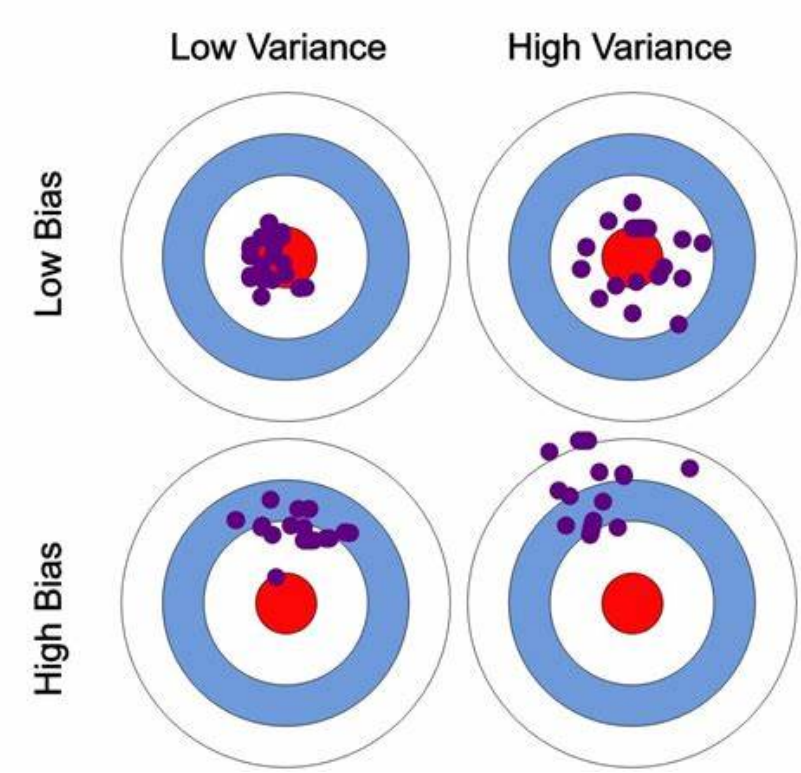
o) **MAE:** Είναι ο μέσος όρος του πραγματικού σφάλματος στις ταξινομήσεις. Γίνεται παίρνοντας την διαφορά των δύο τιμών.

$$Mean\ Absolute\ Error = \frac{1}{n} \times \sum |y_i - \hat{y}_i|$$

p) **MAPE:** Είναι ο ποσοστιαίος μέσος όρος του πραγματικού σφάλματος στις ταξινομήσεις. Γίνεται ποσοστιαία παίρνοντας την διαφορά των δύο τιμών. Πρέπει να χρησιμοποιείται σε μεγάλα σύνολα δεδομένων διότι αν οι τιμές είναι δυο πχ. 2 και 1 θα έχουμε 2-1/2=0.5 άρα 50% σφάλμα κάτι που δείχνει η πρόβλεψη είναι αρκετά μακριά ενώ είναι μια μονάδα.

$$Mean\ Absolute\ Prediction\ Error = 100\% \times (\frac{1}{n} \times \sum |y_i - \hat{y}_i|)$$

Παρακάτω θα αναλύσουμε κάποιες επίσης χρήσιμες ορολογίες [28]:



Εικόνα 13. Bias and Variance

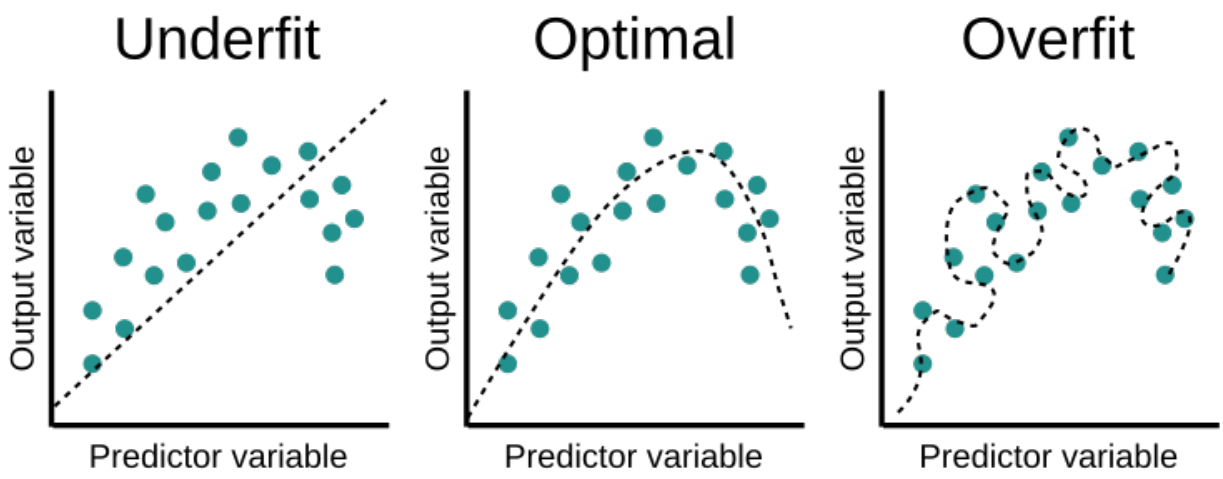
Πηγή: https://nvsyashwanth.github.io/machinelearningmaster/bias-variance/

Bias: Το bias που θα μπορούσαμε σε μια ελεύθερη μετάφραση να το πούμε μεροληψία, αναφέρεται στο σφάλμα εξαιτίας πολύ απλών υποθέσεων στον αλγόριθμο. Αυτές οι υποθέσεις κάνουν το μοντέλο πιο εύκολο να κατανοηθεί και να μάθει αλλά μπορεί να μην καταλάβει την πολυπλοκότητα των δεδομένων. Είναι το σφάλμα εξαιτίας της ανικανότητας του υπολογιστή να αναπαραστήσει την πραγματική σχέση μεταξύ των εισόδων και των εξόδων με ακρίβεια. Όταν ένα μοντέλο έχει κακή απόδοση και στην εκπαίδευση και στην αξιολόγηση τότε συνεπάγεται σε υψηλό bias εξαιτίας του απλού μοντέλου, και αυτό δείχνει ότι υπάρχει underfitting το οποίο θα αναλυθεί παρακάτω.

Διακύμανση (Variance): Αντιθέτως η Διακύμανση είναι το σφάλμα εξαιτίας της ευαισθησίας του μοντέλου στις διακυμάνσεις στα δεδομένα εκπαίδευσης. Είναι το εύρος μεταβλητότητας των προβλέψεων του μοντέλου για διάφορα στιγμιότυπα των δεδομένων εκπαίδευσης. Υψηλή διακύμανση συμβαίνει όταν το μοντέλο μαθαίνει τον θόρυβο των δεδομένων και τυχαίες διακυμάνσεις αντί για το μοτίβο που κρύβεται ανάμεσα. Σαν αποτέλεσμα το μοντέλο έχει καλή απόδοση στα δεδομένα εκπαίδευσης αλλά υπολειτουργεί στα δεδομένα αξιολόγησης, και αυτό δείχνει overfitting το οποίο θα αναλυθεί στην συνέχεια.

Υπερπροσαρμογή (Overfitting): Ένα μοντέλο λέγεται ότι είναι υπερπροσαρμοσμένο όταν δεν κάνει ακριβής προβλέψεις στα δεδομένα αξιολόγησης. Όταν ένα μοντέλο εκπαιδεύεται με τόσα πολλά δεδομένα , αρχίζει να μαθαίνει από τον θόρυβο και τα μη ακριβής δεδομένα εισόδου στο σύνολο δεδομένων μας. Και όταν αξιολογούμε με τα δεδομένα αξιολόγησης τα αποτελέσματά μας είναι υψηλής διακύμανσης. Τότε το μοντέλο δεν κατηγοριοποιεί τα δεδομένα σωστά, εξαιτίας πολλών λεπτομερειών και θορύβου. Συνοψίζοντας, Υπερπροσαρμογή είναι όταν η αξιολόγηση των αλγορίθμων μηχανικής μάθησης στα δεδομένα εκπαίδευσης είναι διαφορετικά από τα δεδομένα που το μοντέλο δεν έχει ξαναδεί.

Υποπροσαρμογή (Underfitting): Ένα μοντέλο λέγεται ότι είναι υποπροσαρμοσμένο όταν είναι πολύ απλό για να εντοπίσει πολύπλοκα δεδομένα. Εκπροσωπεί την ανικανότητα του μοντέλου να μάθει τα δεδομένα εκπαίδευσης αποτελεσματικά με αποτέλεσμα την κακή επίδοση και στα δεδομένα εκπαίδευσης και αξιολόγησης. Με απλούς όρους ένα υποπροσαρμοσμένο μοντέλο είναι ανακριβής όταν δει καινούρια δεδομένα. Κυρίως συμβαίνει όταν χρησιμοποιούμαι ένα πολύ απλό μοντέλο με εξαιρετικά απλοποιημένες προβλέψεις. Για να αντιμετωπίσουμε αυτό το πρόβλημα πρέπει να χρησιμοποιήσουμε πιο πολύπλοκα μοντέλα με ενισχυμένη αναπαράσταση χαρακτηριστικών και λιγότερη κανονικοποίηση.



Εικόνα 14. Overfitting/Underfitting

Πηγή: https://www.educative.io/answers/overfitting-and-underfitting

4.2 Τύποι Μηχανικής Μάθησης

4.2.1 Επιβλεπόμενη μάθηση

Η επιβλεπόμενη μάθηση (Supervised learning) αποτελεί έναν πυλώνα στην ανάπτυξη τεχνητής νοημοσύνης και μηχανικής μάθησης. Πρόκειται για μια μεθοδολογία όπου ένα μοντέλο τεχνητής νοημοσύνης εκπαιδεύεται χρησιμοποιώντας ένα σύνολο δεδομένων τα οποία έχουν ήδη καταταγμένα σε κατηγορίες (ή έχουν συγκεκριμένες τιμές στην περίπτωση της παλινδρόμησης). Αυτά τα δεδομένα χρησιμοποιούνται για να διδάξουν στο μοντέλο πώς να αναγνωρίζει τα πρότυπα και τις σχέσεις που προσδιορίζουν την κατηγορία ή την τιμή ενός νέου δείγματος δεδομένων. Αυτή η μέθοδος είναι ιδιαίτερα δημοφιλής διότι επιτρέπει στο μοντέλο να κάνει προβλέψεις ή να λάβει αποφάσεις με βάση την εμπειρία που έχει αποκτήσει κατά την εκπαίδευση. Για παράδειγμα, ένα μοντέλο επιβλεπόμενης μάθησης μπορεί να εκπαιδευτεί για να διαχωρίζει τα email σε spam ή όχι με βάση τα χαρακτηριστικά του κειμένου τους. Επίσης, μπορεί να εκπαιδευτεί για να προβλέψει την τιμή μετοχών, την απόδοση ενός φοιτητή σε εξετάσεις ή ακόμα και το προσδόκιμο ζωής ενός ασθενούς με βάση ιατρικά δεδομένα [29].

Η διαδικασία της επιβλεπόμενης μάθησης συχνά περιλαμβάνει τη χρήση ενός σετ δεδομένων για εκπαίδευση και ένα ξεχωριστό σετ για δοκιμή του μοντέλου. Το σετ εκπαίδευσης χρησιμοποιείται για να διδάξει στο μοντέλο τη σωστή αντιστοίχιση μεταξύ εισόδων και εξόδων, ενώ το σετ δοκιμής επιτρέπει στους ερευνητές να αξιολογήσουν πόσο καλά το μοντέλο γενικεύει αυτές τις αντιστοιχίσεις σε νέα δεδομένα.

Η επιβλεπόμενη μάθηση είναι επίσης γνωστή για την ικανότητα της να διαχειρίζεται και να λύνει ποικίλα προβλήματα ταξινόμησης και παλινδρόμησης. Στην ταξινόμηση, ο στόχος είναι να κατηγοριοποιήσει τα δεδομένα σε προκαθορισμένες κλάσεις, ενώ στην παλινδρόμηση, το μοντέλο προσπαθεί να προβλέψει έναν συνεχή στόχο. Παραδείγματα περιλαμβάνουν την ταξινόμηση εικόνων, την αναγνώριση ομιλίας, την πρόβλεψη καιρού, ή την εκτίμηση κινδύνων σε ασφαλιστικά προϊόντα.

Επιπλέον, η επιβλεπόμενη μάθηση παίζει βασικό ρόλο στη βελτίωση της αυτοματοποίησης και της λήψης αποφάσεων σε πολλούς τομείς, από την ιατρική διάγνωση μέχρι την ανάλυση της αγοράς και τη διαχείριση των πόρων. Με την αυξανόμενη διαθεσιμότητα των μεγάλων δεδομένων και την πρόοδο στην υπολογιστική ισχύ, η επιβλεπόμενη μάθηση αναμένεται να συνεχίσει να παίζει έναν πρωταγωνιστικό ρόλο στην εξέλιξη της τεχνητής νοημοσύνης.

4.2.2 Μη επιβλεπόμενη μάθηση

Η μη επιβλεπόμενη μάθηση αντιπροσωπεύει ένα διαφορετικό παράδειγμα στη μηχανική μάθηση σε σύγκριση με την επιβλεπόμενη μάθηση. Αντί να χρησιμοποιεί επισημειωμένα δεδομένα, αυτή η προσέγγιση αναζητά κρυμμένα μοτίβα ή δομές σε μη επισημειωμένα δεδομένα. Ο βασικός στόχος της μη επιβλεπόμενης μάθησης είναι η ανακάλυψη νέων μοτίβων ή η ομαδοποίηση των δεδομένων σε υποομάδες χωρίς προκαθορισμένες ετικέτες ή κατηγορίες [30]. Ένας από τους κύριους τομείς όπου η μη επιβλεπόμενη μάθηση έχει αναπτυχθεί είναι η επεξεργασία μοριακών και ατομικών δεδομένων προσομοίωσης στις επιστήμες υλικών, στη φυσική στερεάς κατάστασης, στη βιοφυσική και στη βιοχημεία. Εδώ χρησιμοποιείται για την ανάλυση μεγάλων ποσοτήτων δεδομένων που παράγονται, αναδεικνύοντας τη σημασία της στην ανίχνευση και στην κατανόηση μη αναμενόμενων προτύπων και δομών [31]. Επίσης, η μη επιβλεπόμενη μάθηση έχει εισχωρήσει σε μια ευρεία γκάμα θεμάτων όπως η αστική μελέτη και ο σχεδιασμός, όπου προσφέρει νέες ερευνητικές ευκαιρίες και προκλήσεις. Αυτή η προσέγγιση παρέχει στατιστικές εισόδους στην εξέλιξη και τις επικρατούσες τάσεις, ανοίγοντας το δρόμο για την αυτοματοποίηση και τη γενίκευση των μεθόδων μηχανικής μάθησης.

Οι κύριες μέθοδοι μη επιβλεπόμενης μάθησης περιλαμβάνουν τη μείωση διάστασης, τη συσταδοποίηση και τις μεθόδους βασισμένες στη βαθιά μάθηση. Η μείωση διάστασης εστιάζεται στην απλοποίηση των δεδομένων χωρίς να χάνεται σημαντική πληροφορία, ενώ η συσταδοποίηση ομαδοποιεί τα δεδομένα με βάση την ομοιότητα μεταξύ τους. Συνολικά, η μη επιβλεπόμενη μάθηση παρέχει μια πιο ευέλικτη και αυτοματοποιημένη προσέγγιση στη μηχανική μάθηση, αφού απαλλάσσει την ανάγκη για επισημειωμένα δεδομένα και χειροκίνητη μηχανική χαρακτηριστικών, επιτρέποντας τη γενίκευση και την εφαρμογή σε μια πληθώρα περιστάσεων [32].

4.2.3 Ενισχυτική μάθηση

Η ενισχυτική μάθηση είναι άλλο ένα παρακλάδι της μηχανικής μάθησης αλλά διαφέρει από τα προηγούμενα. Σε γενικές γραμμές αυτό που κάνει είναι να συνεργάζεται με το περιβάλλον του για να αυξήσει τις ανταμοιβές του πράττοντας και μαθαίνοντας από τις παρενέργειες. Για να κάνει το μοντέλο να πράξει χρησιμοποιεί δύο μεθόδους, εκμετάλλευση, όπου το μοντέλο μέσο της εμπειρίας του από τα μονοπάτια και τα λάθη που έχει κάνει διαλέγει τις καινούριες ενέργειες του, και η δεύτερη μέθοδος είναι η εξερεύνηση που δεν βασίζεται στην προηγούμενη εμπειρία του. Η ενισχυτική μάθηση είναι σαν κάποιον που είναι αυτοδίδακτος που κανένας δεν τον βοηθάει στο τι να κάνει. Έτσι το μοντέλο θα κάνει μόνο του τις ενέργειες του έτσι ώστε να φτάσει στο επιθυμητό αποτέλεσμα. Κάτι άλλο ενδιαφέρον για αυτού του είδους μάθηση είναι πως στόχος της είναι το βασικό πρόβλημα της τεχνητής νοημοσύνης, που είναι η εμφάνιση της πληροφορίας. Χρησιμοποιείται για πολλούς σκοπούς μερικοί από αυτούς είναι [33]:

- **Ρομποτική:** Μπορεί να χρησιμοποιηθεί στην ρομποτική για διάφορες εργασίες όπως το περπάτημα και άλλες πιο περίπλοκες εργασίες. Χάρη στον μηχανισμό trial and error που χρησιμοποιεί είναι εφαρμόσιμο σε πολλές τέτοιες εργασίες.
- **Βιντεοπαιχνίδια:** Βρίσκουν τρόπους να αναβαθμίζουν το σκορ τους προσπερνώντας ακόμα και την ανθρώπινη απόδοση.
- **Συστήματα προτάσεων:** Όπως την υπηρεσία Νέτφλιξ που γίνεται πρόταση στον χρήστη για ταινίες που μπορεί να τον ενδιέφεραν σύμφωνα με υλικό που έχει δει παλιότερα, αυτό είναι άλλη μια από τις δυνατότητες της ενισχυτικής μάθησης.
- **Αυτόνομα αυτοκίνητα:** Χρησιμοποιείται για την ανάπτυξη συστημάτων ελέγχου, δίνοντας τους την δυνατότητα να προσαρμόζονται σε δυναμικό και απρόβλεπτο περιβάλλον.
- **Συστήματα ενέργειας:** Ακόμα και στα Smart grids που αναφέραμε στο προηγούμενο κεφάλαιο μπορεί να χρησιμοποιηθεί για να χειρίζεται την κατανάλωση ενέργειας .

Αυτό το είδος περιλαμβάνει και κάτι που λέγεται αμοιβή [34]. Η αμοιβή είναι ένας μηχανισμός που οδηγεί το μοντέλο της ενισχυτικής μάθησης, αυτό είναι η επιθυμητή τελική κατάσταση που το μοντέλο κυνηγάει. Κάθε φορά που γίνεται μια σωστή κίνηση το μοντέλο δέχεται ένα κομμάτι της αμοιβής, όσο πιο σωστές κινήσεις τόσο μεγαλύτερη η αμοιβή. Για παράδειγμα, σε μια παρτίδα σκάκι η συνάρτηση αμοιβής θα μπορούσε να δίνει θετική αμοιβή για μια νίκη και μια αρνητική αμοιβή για μια ήττα και μπορεί μια μικρότερη αμοιβή για κάθε σωστό βήμα και μικρή τιμωρία για κάθε λάθος βήμα αντίστοιχα. Το πως θα λειτουργεί η συνάρτηση αμοιβής θα έχει μεγάλο αντίτιμο στην απόδοση του μοντέλου.

4.3 Βαθιά Μάθηση

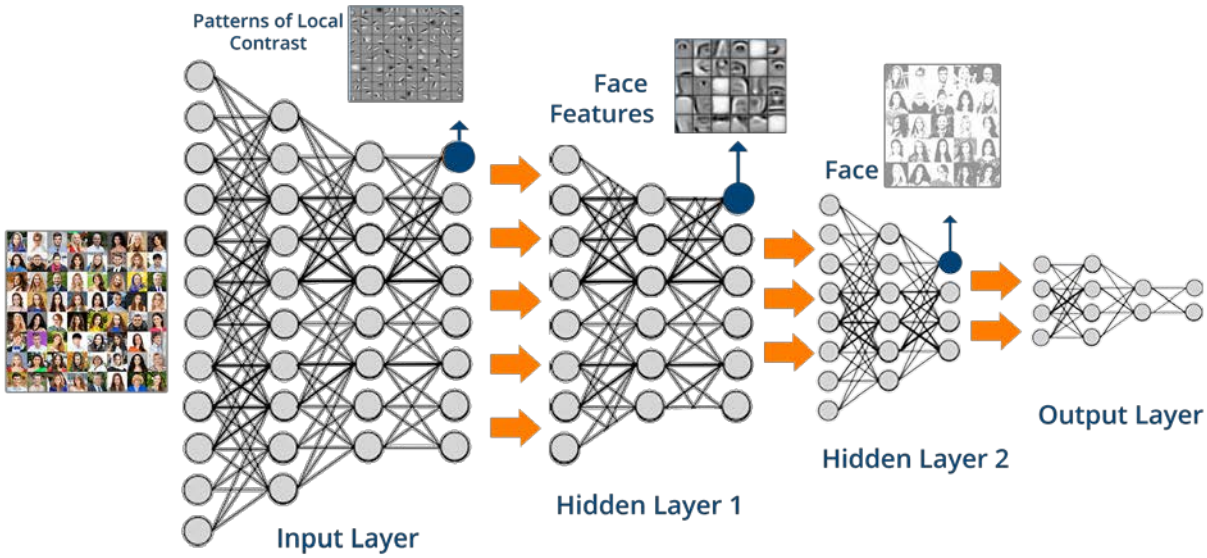
Η τεχνολογία της μηχανικής μάθησης ενεργεί πολλά μέρη της μοντέρνας κοινωνίας: από αναζητήσεις του ίντερνετ μέχρι φιλτράρισμα των περιεχομένων στα μέσα κοινωνικής δικτύωσης μέχρι σε προτάσεις σε σελίδες προβολής ταινιών, και ακόμα περισσότερο σε προϊόντα καταναλωτών όπως κάμερες και κινητά τηλέφωνα. Χρησιμοποιούνται για την αναγνώριση πραγμάτων σε εικόνες, μεταγραφή της ομιλίας σε κείμενο, αντιστοίχιση ειδήσεων, αναρτήσεων ή προϊόντων με τα ενδιαφέροντα των χρηστών και επιλογή σχετικών αποτελεσμάτων αναζήτησης. Αυξανόμενα αυτές οι εφαρμογές χρησιμοποιούν μια πληθώρα από τεχνικές, αυτές οι τεχνικές ονομάζονται βαθιά μάθηση (Deep Learning).

Οι γνωστές τεχνικές μηχανικής μάθησης ήταν περιορισμένες ως προς την ικανότητά τους να επεξεργάζονται φυσικά δεδομένα σε ωμή μορφή. Για πολύ καιρό, η δημιουργία αναγνώρισης μοτίβο ή συστημάτων μηχανικής μάθησης χρειαζόταν προσεκτική μηχανική και έναν μεγάλο βαθμό γνώσης του τομέα για να φτιαχτεί ένας αλγόριθμος εξαγωγής χαρακτηριστικών που μετάτρεπε τα ωμά δεδομένα σε χρήσιμές πληροφορίες που το μοντέλο μπορούσε να χρησιμοποιήσει για να ανακαλύψει μοτίβο στα εισαγόμενα δεδομένα.

Η μάθηση με αναπαράσταση [35] είναι ένα σύνολο από μεθόδους που επιτρέπουν σε μια μηχανή να τραφεί με ωμά δεδομένα και αυτόματα να ανακαλύψει τις αναπαραστάσεις που χρειάζονται για αναγνώριση ή ταξινόμηση. Η μέθοδος της βαθιάς μάθησης είναι η μάθηση με αναπαράσταση, που αναφέραμε προηγουμένως, με πολλαπλά επίπεδα αναπαραστάσεων, που επιτυγχάνονται με τη σύνθεση απλών αλλά μη γραμμικών μονάδων που μετατρέπουν κάθε φορά την αναπαράσταση σε ένα επίπεδο (ξεκινώντας από την εισαγωγή ωμών δεδομένων) σε μια αναπαράσταση σε ένα υψηλότερο κάπως πιο γενικό επίπεδο. Με τον συνδυασμό αρκετών τέτοιων μεταμορφώσεων , το μοντέλο μπορεί να μάθει πολύ περίπλοκους συνδυασμούς.

Όσον αφορά τις εργασίες ταξινόμησης, ένα υψηλότερο επίπεδο αναπαραστάσεων μεγιστοποιούν πληροφορίες που είναι χρήσιμες για την διάκριση των κλάσεων και ελαχιστοποιούν τις αχρείαστες παραλλαγές. Για παράδειγμα μια εικόνα παράγεται υπό την μορφή ενός πίνακα τιμών πίξελ, και τα χαρακτηριστικά που το μοντέλο μαθαίνει στο πρώτο επίπεδο αναπαράστασης τυπικά αναπαριστούν την παρουσία ή την απουσία άκρων σε συγκεκριμένα μέρη και κατευθύνσεις της εικόνας. Το δεύτερο επίπεδο συνήθως εντοπίζει μοτίβο παρατηρώντας συγκεκριμένες διατάξεις των άκρων χωρίς να έχει σημασία η μικρή παραλλαγή στις τοποθεσίες των άκρων. Το τρίτο επίπεδο μπορεί να συγκεντρώσει κάποια μοτίβο σε μεγαλύτερα συνδυαστικά κομμάτια που αντιστοιχούν σε κομμάτια από γνωστά αντικείμενα. Το κύριο προτέρημα της βαθιάς μάθησης είναι ότι αυτά τα επίπεδα χαρακτηριστικών δεν είναι σχεδιασμένα από μηχανικούς, αντιθέτως έχουν μαθευτεί από δεδομένα χρησιμοποιώντας διαδικασίες μάθησης γενικής χρήσης.

Η βαθιά μάθηση κάνει μεγάλα βήματα στη λύση προβλημάτων που έχουν αντέξει για πολλά χρόνια τις καλύτερες προσπάθειες τη τεχνίτης νοημοσύνης. Έχει αποδειχθεί εξαιρετικό στην ανακάλυψη πολύπλοκων δομών σε δεδομένα πολλών διαστάσεων, έτσι είναι εφαρμόσιμο σε πολλές πτυχές της επιστήμης, επιχείρησης και κυβέρνησης. Είναι πολύ πιθανό η βαθιά μάθηση να έχει πολλές ακόμα επιτυχίες στο κοντινό μέλλον επειδή χρειάζεται ελάχιστη μηχανική από ανθρώπους έτσι μπορεί να εκμεταλλευτεί τα αυξανόμενα παγκόσμια δεδομένα. Καινούριοι αλγόριθμοι και αρχιτεκτονικές που παράγονται αυτή τη στιγμή, το μόνο που θα καταφέρουν είναι την ταχύτερη εξέλιξη της βαθιάς μάθησης.



Εικόνα 15. Επίπεδα Deep Learning

Πηγή: https://thedatascientist.com/what-deep-learning-is-and-isnt/

5. Αλγόριθμοι

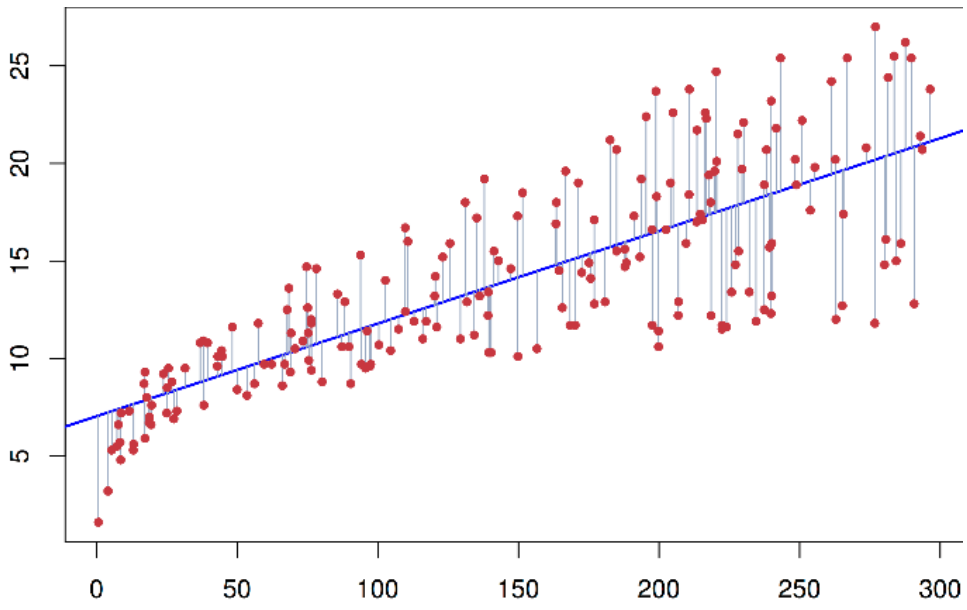
5.1 Αλγόριθμοι Μηχανικής Μάθησης

5.1.1 Linear Regression

Linear Regression ή αλλιώς Γραμμική Παλινδρόμηση [36] είναι ένας από τους πιο βασικούς αλγορίθμους στον τομέα της μηχανικής μάθησης και χρησιμοποιείται για την πρόβλεψη συνεχόμενων στοχευμένων μεταβλητών, βασισμένων σε ένα η περισσότερα χαρακτηριστικά που έχουν δοθεί στην είσοδο. Είναι μια μαθηματική τεχνική για την εισαγωγή συγκεκριμένων συνόλων δεδομένων σε μια συνάρτηση, τυπικά σε μια ευθεία γραμμή για αυτό το λόγο λέγεται γραμμική. Η βασική ιδεολογία αυτού του αλγορίθμου είναι να βρεθεί η ευθεία γραμμή που ταιριάζει καλύτερα στο γράφημα και μπορεί να προβλέψει όσο πιο ακριβέστατα γίνεται τις τιμές εξόδου μέσα σε ένα αποδεκτό εύρος. Σε ένα απλό μοντέλο γραμμικής παλινδρόμησης υπάρχει συνήθως μια μεταβλητή η οποία είναι ανεξάρτητη και μια η οποία είναι εξαρτημένη. Η σχέση που έχουν αυτές οι μεταβλητές δίνεται από την παρακάτω εξίσωση.

$$Y = \alpha + \beta X + \in$$

Όπου Y είναι η εξαρτώμενη μεταβλητή, X είναι η ανεξάρτητη μεταβλητή, α και β είναι οι παράμετροι του μοντέλου παλινδρόμησης, και $\in$ είναι το σφάλμα. Σε μερικές περιπτώσεις που υπάρχουν παραπάνω από μια ανεξάρτητη μεταβλητή ονομάζεται πολλαπλή γραμμική παλινδρόμηση.



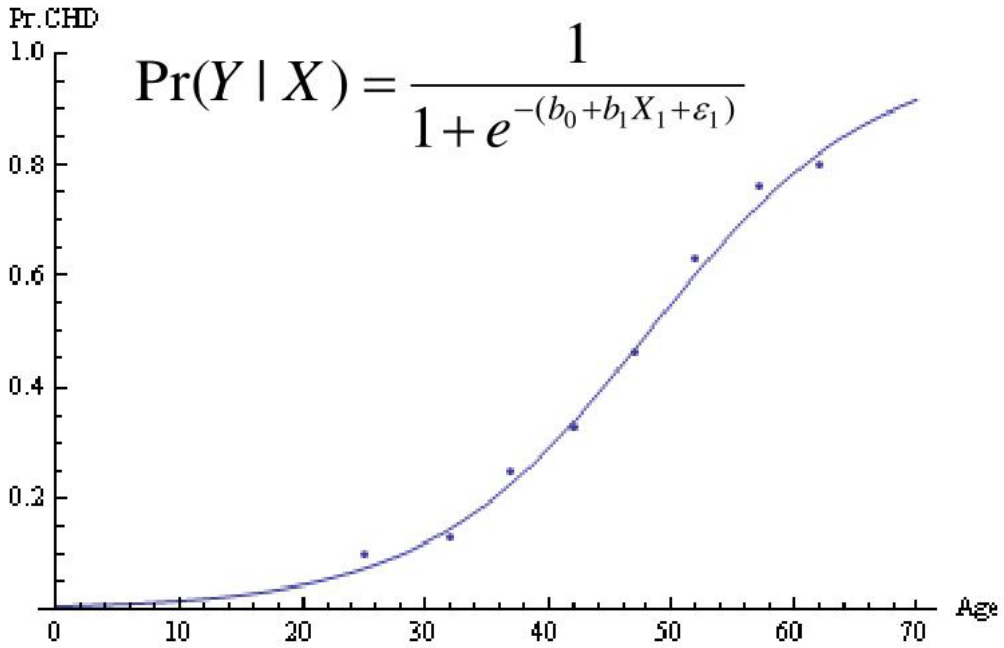
Εικόνα 16. Linear Regression

Πηγή: https://www.hcbravo.org/IntroDataSci/bookdown-notes/linear-regression.html

5.1.2 Logistic Regression

Logistic Regression ή αλλιώς Λογιστική Παλινδρόμηση είναι ένας δυνατός αλγόριθμος ο οποίος συνήθως χρησιμοποιείται για τον υπολογισμό της πιθανότητας ενός δυαδικού αποτελέσματος, βασισμένο σε μια ή περισσότερες ανεξάρτητες μεταβλητές. Σε αντίθεση με την Γραμμική Παλινδρόμηση η οποία προβλέπει συνεχόμενα αποτελέσματα, η Λογιστική Παλινδρόμηση με δυαδικά ή διχοτομημένα αποτελέσματα κάτι που σημαίνει πως μπορούν να έχουν μόνο δύο τιμές εξόδου όπως 0 και 1, ναι και όχι, αληθές ψευδές. Είναι ένας αλγόριθμος ο οποίος βοηθάει σε κατηγορίες που επίσης τα δεδομένα του ανήκουν σε μια από τις δύο κατηγορίες. Η λογιστική παλινδρόμηση υπολογίζει τον λογάριθμο της πιθανότητας ενός συμβάντος να πραγματοποιηθεί, για αυτόν τον λόγο είναι γνωστό στον τομέα της υγείας για υπολογισμούς πιθανοτήτων μιας ασθένειας να συμβεί, η και στα οικονομικά για τα σκορ χρεωστικών καρτών.

Ο μηχανισμός πίσω από την λογιστική παλινδρόμηση αντιπροσωπεύει αυτόν της γραμμικής παλινδρόμησης αλλά με ενσωματωμένη μια λογιστική συνάρτηση για να αποσπάσει την πιθανότητα ενός δυαδικού αποτελέσματος. Η λογιστική συνάρτηση διαβεβαιώνει ότι η υπολογισμένη πιθανότητα είναι εγκλωβισμένη ανάμεσα στο 0 και το 1. Οι παράμετροι της λογιστικής παλινδρόμησης υπολογίζονται με την μεγιστοποίηση της πιθανότητας παρατήρησης των τιμών του δείγματος και όχι με την ελαχιστοποίηση του αθροίσματος των τετραγωνικών σφαλμάτων όπως στην γραμμική παλινδρόμηση [37].



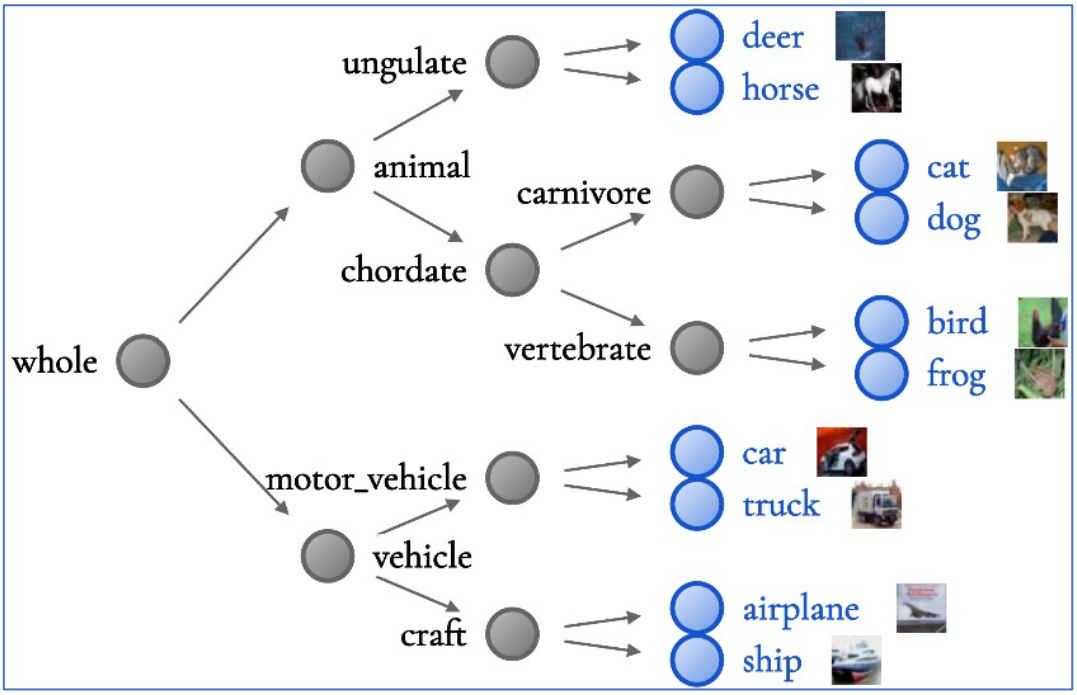
Εικόνα 17. Logistic Regression

Πηγή: https://www.slideserve.com/aristotle-graves/logistic-regression-analysis

5.1.3 Decision Trees

Τα Decision Trees ή αλλιώς Δέντρα Αποφάσεων [38] είναι γραφικές αναπαραστάσεις για την δημιουργία μοντέλων ταξινόμησης χωρίζοντας τα δεδομένα σε υποκατηγορίες βασιζόμενα στις εισαγόμενες τιμές τους, δημιουργώντας έτσι ένα διάγραμμα ροής. Τα μοντέλα που δημιουργούνται παριστάνουν την λογική ενός ανθρώπου και είναι εύκολα κατανοητά, τα οποία τα κάνει δημοφιλή και σε πολλούς άλλους τομείς. Η δομή αυτών περιέχει κόμβους οι οποίοι παριστάνουν αξιολογήσεις στα χαρακτηριστικά, κλαδιά να ανταπαντούν στα αποτελέσματα αυτών των αξιολογήσεων και φύλα κόμβοι που δείχνουν την πρόβλεψη στην έξοδο ή ετικέτα κλάσης. Τα μοντέλα δένδρου απόφασης, μπορούν να δεχθούν τόσο αριθμητικές όσο και κατηγορηματικές εισόδους δεδομένων και η διαφάνεια τους οδηγεί σε ένα ερμηνεύσιμο αποτέλεσμα, το οποίο τα κάνει ένα πολύτιμο εργαλείο σε διαδικασίες λήψεις αποφάσεων και σε πολλούς άλλους επιστημονικούς κλάδους.

Με την πάροδο τον χρόνων τα μοντέλα αυτά έχουν αποδείξει την χρησιμότητα τους στην πληθώρα των περιπτώσεων λόγο της εύκολης ερμηνεύσεις τους καθώς και της στιβαρότητας τους ακόμα και σε περιπτώσεις με ελλειμματικές εισόδους. Θεωρούνται αποτελεσματικές μέθοδοι για την εξόρυξη δεδομένων και έχουν υιοθετηθεί ευρέως σε πολλούς κλάδους. Αξιοσημείωτοι αλγόριθμοι για την δημιουργία μοντέλων δέντρου απόφασης, αποτελούν οι C4.5 και CART (Classification and Regression Trees) μεταξύ πολλών άλλων. Οι αλγόριθμοι αυτοί αναλύουν ένα τμήμα εκπαιδευτικών παραδειγμάτων, με γνωστό αποτέλεσμα, για την δημιουργία ενός δέντρου διαλέγοντας τα καλυτέρα χαρακτηριστικά για κάθε κόμβο για να επιτευχθούν τα πιο ομοιογενή υποσύνολα βελτιστοποιώντας έτσι την ταξινόμηση καθώς και την παλινδρόμηση των δεδομένων.



Εικόνα 18. Decision Trees

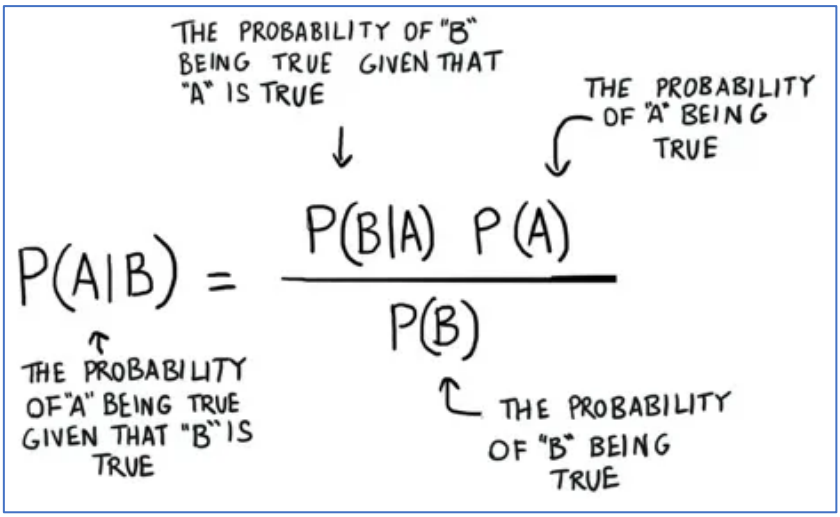
Πηγή: https://bair.berkeley.edu/blog/2020/04/23/decisions/

5.1.4 Naive Bayes

Naive Bayes είναι μια συλλογή από επιβλεπόμενους αλγορίθμους μάθησης που βασίζονται στην εφαρμογή του θεωρήματος του Bayes, με το ‘’αφελείς’’ υποθέτει ότι κάθε ζευγάρι χαρακτηριστικών των δεδομένων είναι υπό κάποιους όρους ανεξάρτητο της τιμής της μεταβλητής της κλάσης. Αυτή η υπόθεση επιτρέπει την δημιουργία μιας στιβαρής καθώς και εύκολης μεθοδολογίας ταξινόμησης. Οι μέθοδοι Naive Bayes προαναγγέλλονται για την απλότητα και την αποτελεσματικότητά τους στην προγνωστική μοντελοποίηση. Το μοντέλο περιλαμβάνει δυο πρωταρχικούς τύπους πιθανοτήτων που προέρχονται κατευθείαν από τα δεδομένα της εκπαίδευσης:

- Την πιθανότητα κάθε κλάσης και
- Την υπό όρους πιθανότητα λαμβάνοντας υπόψη τις τιμές όλων των χαρακτηριστικών.

Ο πυρήνας του Naive Bayes [39] εξελίσσεται γύρω από το θεώρημα Bayes το οποίο χρησιμοποιείτε για τον υπολογισμό των μεταγενέστερων πιθανοτήτων κάθε κλάσης, δεδομένου των χαρακτηριστικών και στη συνέχεια να επιλέξει την κλάση με την υψηλότερη μεταγενέστερη πιθανότητα ως την προβλεπόμενη κλάση. Παρά την υπόθεση της ανεξαρτησίας μεταξύ των χαρακτηριστικών της κλάσης η οποία συχνά παραβιάζεται, ο Naive Bayes τακτικά παρέχει ανταγωνιστικά ακριβή ταξινόμηση καθιστώντας το μια αξιόπιστη επιλογή σε διαφορά σενάρια ταξινόμησης της καθημερινότητας. Η μαθηματική του διατύπωση επιτρέπει μια σειρά πιθανολογικών υπολογισμών στοιχειωμένες στην εύρεση της καλύτερης προσαρμοζόμενης ταξινόμησης ενός τμήματος δεδομένων με συγκεκριμένο πρόβλημα. Η ευκολία εφαρμογής του, συνδυαζόμενη με την ικανότητα να χειρίζεται έναν μεγάλο αριθμό χαρακτηριστικών και η αποτύπωση πιθανολογικών προβλέψεων κάνει τον αλγόριθμο αυτό έναν ευρέως υιοθετημένο αλγόριθμο σε διάφορους κλάδους όπως στην ταξινόμηση κείμενων, στο φιλτράρισμα ανεπιθύμητων μνημάτων καθώς και στην ανάλυση συναισθημάτων.

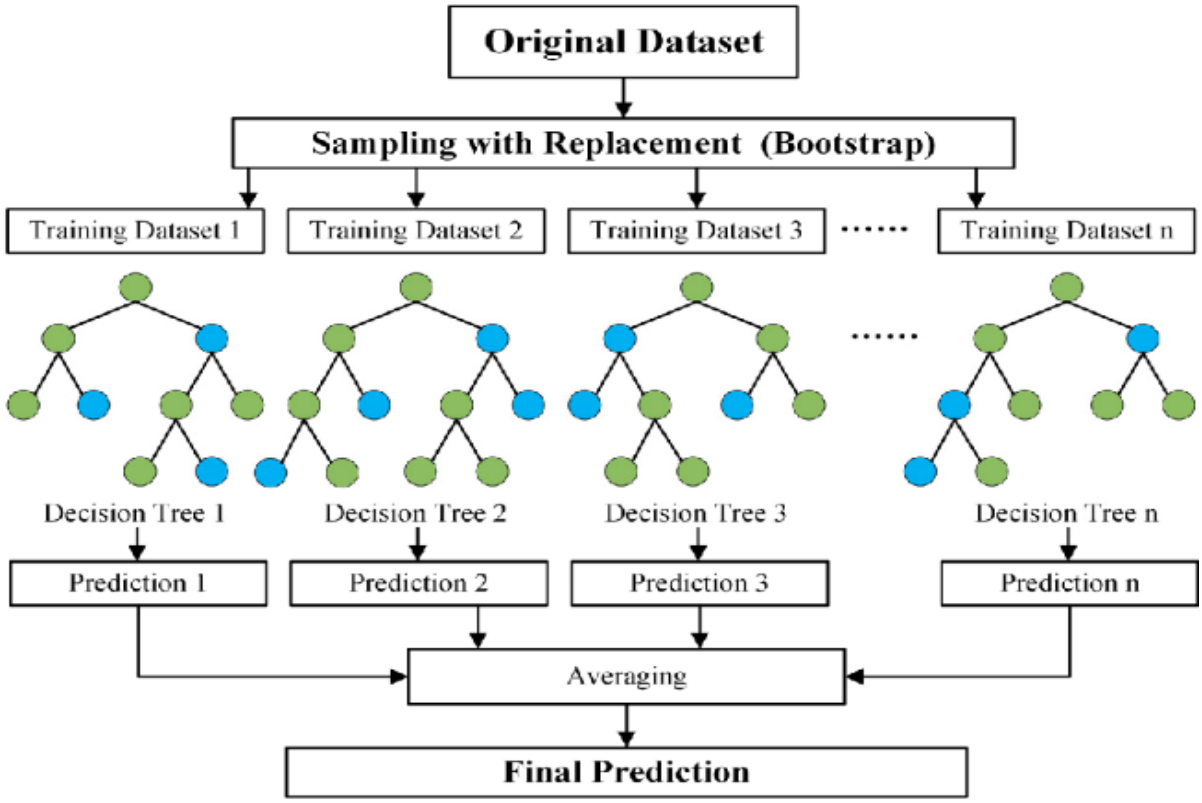


Εικόνα 19. Naive Bayes

5.1.5 Random Forests

Τα Random Forests ή αλλιώς Τυχαία Δάση λειτουργούν χτίζοντας πολλαπλά δέντρα αποφάσεων κατά την διάρκεια της εκπαίδευσης και εξάγει την κλάση που είναι τα mode (η τιμή που εμφανίζεται πιο συχνά στο σύνολο των δεδομένων) από τα ατομικά δέντρα για ταξινόμηση ή μέση πρόβλεψη των ατομικών δέντρων για παλινδρομικές εργασίες. Κάθε δέντρο στα τυχαία δάση είναι φτιαγμένο σε ένα υποσύνολο των δεδομένων εκπαίδευσης, ανακατεμένο με το Bootstrap Sample (ένα σύνολο δεδομένων που έχει προβλέψει το κάθε δέντρο) και σε κάθε σημείο ένα τυχαίο υποσύνολο από τα χαρακτηριστικά διαλέγετε για να χωρίσει το σημείο (node). Η τυχαιότητα στην επιλογή δειγμάτων (samples) και χαρακτηριστικών ενώ κατασκευάζονται τα δέντρα διαβεβαιώνει την ποικιλομορφία μεταξύ των δέντρων, τα οποία με την σειρά τους αυξάνουν την σταθερότητα και την γενικότερη απόδοση του μοντέλου [40]. Το σφάλμα της γενίκευσης συγκλίνει όταν τα δέντρα στο δάσος αυξάνονται δημιουργώντας ένα ισχυρό μοντέλο ακόμα και με την παρουσία θορύβου στα δεδομένα.

Αυτός ο αλγόριθμος είναι γνωστός για την μεγάλη του ακρίβεια, την ικανότητα του να χειρίζεται μεγάλο αριθμό χαρακτηριστικών και χαμένες τιμές. Επίσης είναι γνωστός επειδή είναι σχετικά γρήγορος στο να εκπαιδεύεται και να χειρίζεται ανισόρροπα σύνολα δεδομένων (ένας τύπος δεδομένων μεγαλύτερος από τον άλλον). Πολλά πεδία χρησιμοποιούν αυτόν τον αλγόριθμο, μερικά παραδείγματα είναι βιολογία ιατρική, οικονομικά και μάρκετινγκ. Χάρη στην ευκολία χρήσης τους και της διαισθητικής τους φύσης.



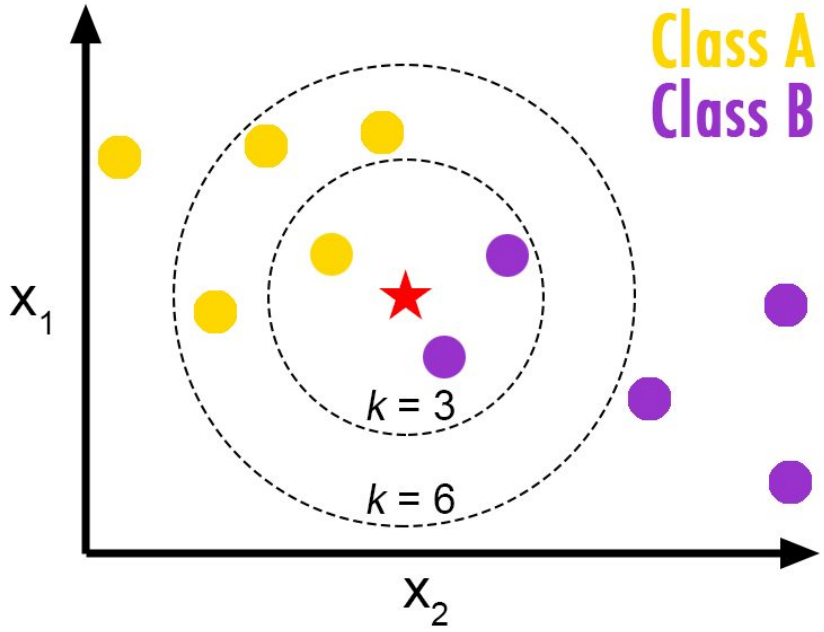
Εικόνα 20. Random Forests

5.1.6 k-Nearest Neighbors (k-NN)

Ο αλγόριθμος k-NN λειτουργεί με κάποια αρχή, ορίζει ένα άγνωστο δείγμα σε μια κλάση ανάλογα με την κλάση με πλειονότητα ανάμεσα στους κοντινούς του γείτονες στον χώρο των χαρακτηριστικών, ή προβλέπει ένα συνεχόμενο αποτέλεσμα βασισμένο στην μέση τιμή των αποτελεσμάτων του πιο κοντινού γείτονα (κ-nearest neighbor). Ο αλγόριθμος υπολογίζει την απόσταση μεταξύ του άγνωστου δείγματος και κάθε άλλου δείγματος στο σύνολο δεδομένων εκπαίδευσης , έπειτα οργανώνει αυτές τις αποστάσεις και τότε διαλέγει τα πιο κοντινά γειτονικά δείγματα για να αποφασίσει το αποτέλεσμα του άγνωστου δείγματος. Το ‘k’ στο k-NN είναι μια παράμετρος που αναφέρεται στον αριθμό των πιο κοντινών γειτόνων για να έχει ο αλγόριθμος υπόψιν όταν κάνει την πρόβλεψη. Η απλότητα του k-NN είναι το ότι δεν έχει παραμέτρους κα είναι πολύ απλό για να το καταλάβει κανείς.

Διάφορες μετρήσεις απόστασης όπως η Ευκλείδεια Απόσταση, Απόσταση Μανχάταν ή Μινκόφσκι Απόσταση μπορούν να χρησιμοποιηθούν στον αλγόριθμο k-NN για να υπολογίσει την ομοιότητα μεταξύ των δειγμάτων. Επιπλέον η αποτελεσματικότητά του δεν περιορίζεται μόνο στην ταξινόμηση, είναι και ένα πολύ χρήσιμο εργαλείο για εργασίες παλινδρόμησης όπου η πρόβλεψη είναι η μέση τιμή των γειτόνων. Μια από τις πιο κρίσιμες αποφάσεις που πρέπει κάποιος να πάρει όταν κατασκευάζει το μοντέλο είναι η τιμή του ‘k’ η οποία επηρεάζει εξαιρετικά την απόδοση του αλγόριθμου. Παρά την απλότητά του, η εξάρτηση του k-NN από τον υπολογισμό των αποστάσεων μεταξύ του

άγνωστου δείγματος και όλων των δειγμάτων του συνόλου δεδομένων εκπαίδευσης μπορεί να γίνει πολύ απαιτητικό σε θέμα υπολογιστικής ισχύς όσο αυξάνεται το μέγεθος του συνόλου δεδομένων, γεγονός που πρέπει να ληφθεί υπόψη κατά την επιλογή του αλγορίθμου k-NN [41].



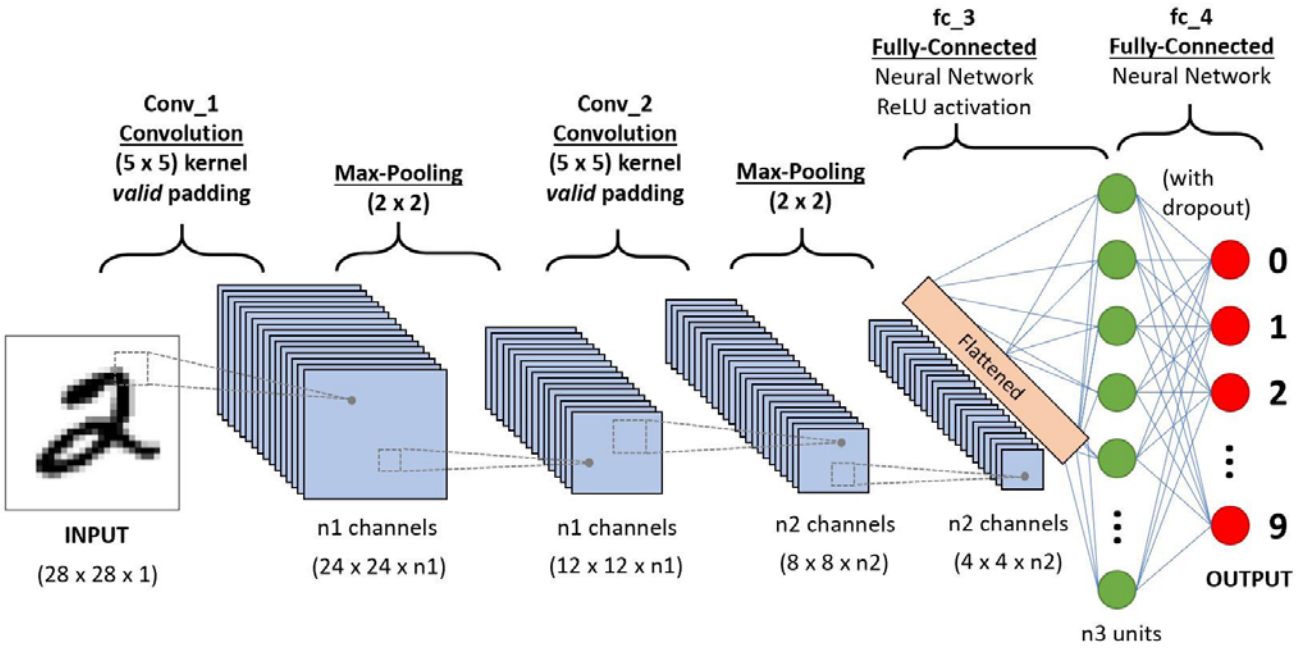
Εικόνα 21. k-NN

Πηγή: https://matlab1.com/knn-classifier/

5.2 Αλγόριθμοι Βαθιάς Μάθησης

5.2.1 Convolutional Neural Network (CNN)

Τα Συνελικτικά Νευρωνικά Δίκτυα [42] έχουν κατασκευαστεί κυρίως για να χειρίζονται εικόνες και είναι εξαιρετικά αποτελεσματικά στο να αναγνωρίζουν διάφορα μοτίβο και χαρακτηριστικά σε εικόνες. Αυτά τα νευρωνικά δίκτυα είναι δομημένα με πολλά επίπεδα, συμπεραλβανομένου τα συνελικτικά επίπεδα , τα επίπεδα συγκέντρωσης και τα πλήρως συνδεδεμένα επίπεδα. Το επίπεδο συνέλιξης εφαρμόζει μια διαδικασία συνέλιξης στην είσοδο, και μεταφέρει το αποτέλεσμα στο επόμενο επίπεδο. Αυτή η διαδικασία επιτρέπει στο δίκτυο να καταλάβει ορισμένα φίλτρα που θα το βοηθάνε για να αναγνωρίσει χαρακτηριστικά σε εικόνες, όπως για παράδειγμα άκρες ή συγκεκριμένες υφές. Το επίπεδο συγκέντρωσης από την άλλη , μειώνει το ύψος και το πλάτος της έντασης εισόδου δεδομένων το οποίο βοηθάει στην μείωση την υπολογιστικής πολυπλοκότητας και απαίτησης και επίσης βοηθάει στην αναγνώριση χαρακτηριστικών ανεξάρτητα στις αλλαγές που γίνονται στο μέγεθος και στην κατεύθυνση. Τέλος, το επίπεδο πλήρους σύνδεσης συνδέει κάθε νευρώνα σε ένα επίπεδο με τον νευρώνα του επομένου επιπέδου, κάτι που κυρίως γίνεται για λόγους ταξινόμησης.



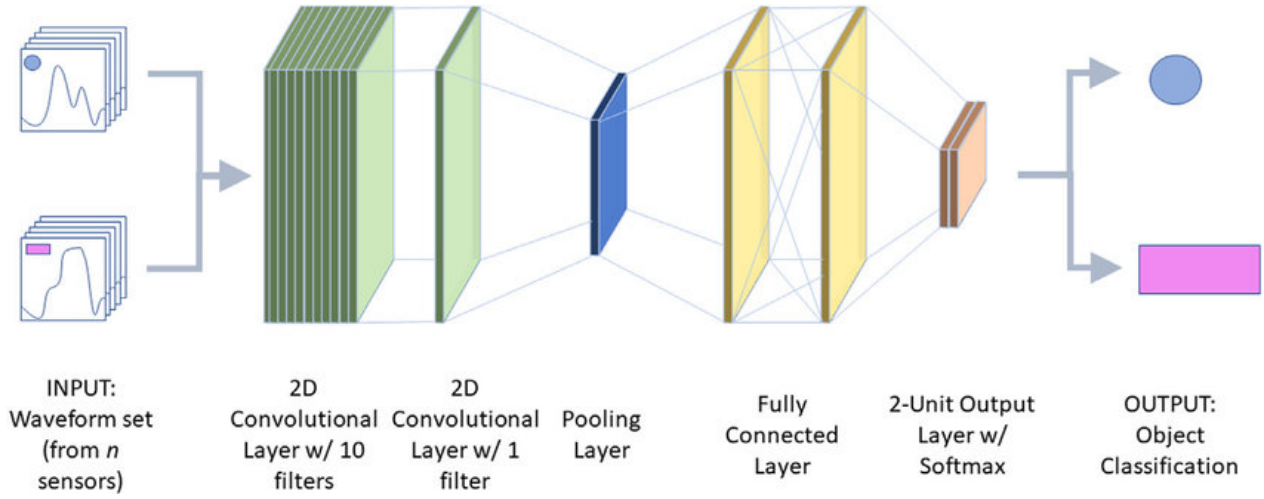
Εικόνα 22. Convolutional Neural Network

Πηγή: https://idiotdeveloper.com/convolution-neural-network-cnn-fundamental-of-deep-learning/

Τα CNN ήταν πολύ χρήσιμα για την ανάπτυξη στον τομέα της όρασης του υπολογιστή, καταφέρνοντας προηγμένες αποδόσεις σε διάφορες εργασίες όπως εικόνα, αναγνώριση βίντεο, συστήματα προτάσεων και επεξεργασία φυσικής γλώσσας. Είναι γνωστά για την ικανότητά τους στο να καταλαβαίνουν αυτόματα και προσαρμοστικά τις χωρικές ιεραρχίες χαρακτηριστικών των εικόνων εισόδου. Αυτό ήταν μια μεγάλη στροφή από τις παραδοσιακές μεθόδους που γινόταν επεξεργασία και γενικότερα μηχανική χαρακτηριστικών. Τα CNN [43] δεν έχουν περιοριστεί μόνο στην αναγνώριση εικόνας, αλλά έχουν δείξει και μεγάλη επιτυχία σε άλλους τομείς όπως ανακάλυψη φαρμάκων, γενετικές εκφράσεις και άλλα στον χώρο της βιο-πληροφορικής.

5.2.2 Temporal Convolutional Network (TCN)

Τα TCN αξιοποιούν μια εξειδικευμένη αρχιτεκτονική για να μάθουν διαφορετικά μοτίβα από ακολουθίες δεδομένων. Τα TCN χρησιμοποιούν συνελικτικά επίπεδα τα οποία τους επιτρέπουν την καταγραφή διάφορων μοτίβων ταυτόχρονα σε διαφορετικά κομμάτια μιας ακολουθίας. Τα ενδιάμεσα επίπεδα στο TCN περιέχουν τόσο μικρά όσο και μεγάλα ιεραρχικά μοτίβα, καθιστώντας τα ικανά να κατανοήσουν τις χρονικές διάφορες σε διάφορα μήκη δεδομένων. Χρησιμοποιούν επίσης διευρυμένες αιτιολογικές συνελίξεις και υπολειμματικά μπλοκ για την εξαγωγή μακροπροθέσμων μοτίβων. Αυτή η αρχιτεκτονική όχι μόνο επιτρέπει στα TCN [44] να εντοπίζει χρονικές διαφορές στα δεδομένα αλλά και να είναι αποτελεσματικά όσον αναφορά τον υπολογιστικό χρόνο, ειδικά για εργασίες που έχουν να κάνουν με πρόβλεψη.



Εικόνα 23. Temporal Convolutional Network

Πηγή:https://www.researchgate.net/figure/Temporal-convolutional-neural-network-architecture_fig5_361482145

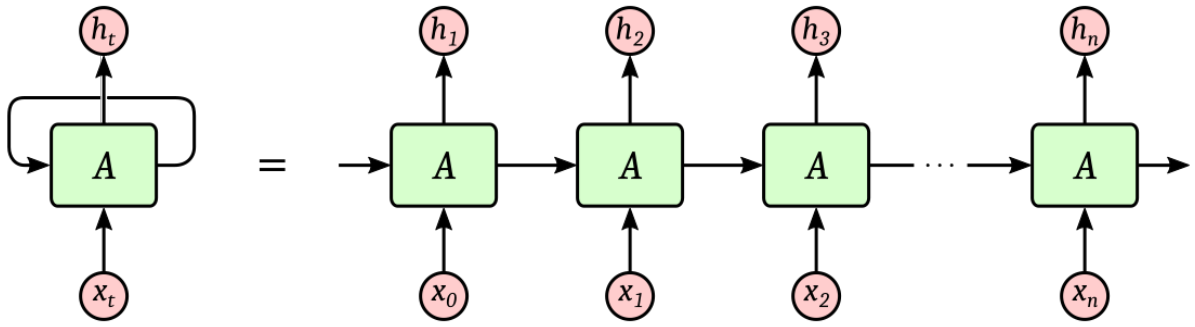
Σε πρακτικές εφαρμογές, τα TCN έχουν χρησιμοποιηθεί για διάφορους σκοπούς , όπως η πρόβλεψη πολυμεταβλητών χρονοσειρών, τμηματοποίηση χρονοσειρών και πολλά άλλα. Η ευελιξία των TCN επιτρέπει την τροποποίηση καθώς και βελτίωση τους, για παράδειγμα, συνδυάζοντας ένα προσωρινό συνελικτικό δίκτυο με μηχανισμό προσοχής καναλιού για τη βελτίωση της απόδοσης στην πρόβλεψη πολυμεταβλητών χρονοσειρών. Επίσης, τα TCN βρίσκουν εφαρμογές μεταξύ πολλών αναδυόμενων πεδίων και έχουν δείξει πλεονεκτήματα σε σχέση με άλλα μοντέλα νευρωνικών δικτυών όταν ασχολούμαστε με δεδομένα ακολουθίας. Η ικανότητα των TCN να χειρίζονται διάφορα μήκη ακολουθιών και να εξάγουν αποτελεσματικά χαρακτηριστικά τα καθιστά ένα πολύτιμο εργαλείο σε ένα ευρύ φάσμα εφαρμογών περιλαμβάνοντας διαδοχικά δεδομένα ή δεδομένα χρονοσειράς [45].

5.2.3 Recurrent Neural Network (RNN)

Τα RNN [46] διαφέρουν θεμελιωδώς από τα παραδοσιακά νευρωνικά δίκτυα τροφοδοσίας, καθώς διαθέτουν μια μορφή μνήμης που τους επιτρέπει να διατηρούν πληροφορίες σε βάθος χρόνου. Αυτό επιτυγχάνεται με τη συμπερίληψη βρόγχων εντός του δικτύου που επιτρέπουν τη μετάδοση πληροφοριών από το ένα βήμα του δικτύου στο επόμενο. Αυτό το μοναδικό χαρακτηριστικό καθιστά τα RNN κατάλληλα για εφαρμογές με ακολουθιακά δεδομένα, καθώς μπορούν να επεξεργαστούν ακολουθίες εισόδων, καθιστώντας τα εξαιρετικά εφαρμόσιμα σε προβλήματα όπως η μοντελοποίηση γλώσσας, η δημιουργία κειμένου και άλλες εργασίες που απαιτούν την κατανόηση της δυναμικής ακολουθίας. Είναι ευρέως γνωστά για την ικανότητα τους να χειρίζονται εφαρμογές με χρονοσειρές και διαδοχικά δεδομένα.

Σε πρακτικές εφαρμογές, τα RNN έχουν αποδειχθεί να είναι αποτελεσματικά σε διάφορους τομείς. Για παράδειγμα, χρησιμοποιούνται συχνά για μεταφράσεις, επεξεργασία φυσικής γλώσσας (NLP), αναγνώριση ομιλίας και σχολιασμό εικόνων. Ενσωματώνονται επίσης, σε

γνωστές εφαρμογές όπως η Siri, για φωνητικές αναζητήσεις και στο Google αναδεικνύοντας έτσι την ικανότητα τους να χειρίζονται αποτελεσματικά τακτικά ή χρονικά προβλήματα. Ορισμένες προηγμένες παραλλαγές των RNN [47], όπως οι μονάδες μακράς βραχυπρόθεσμης μνήμης (LSTM) και ο μηχανισμός προσοχής, έχουν αναπτυχθεί για την αντιμετώπιση προκλήσεων όπως the vanishing gradient problem, αναδεικνύοντας περεταίρω της δυνατότητα να χειρίζονται εξαρτήσεις μεγάλης εμβέλειας σε ακολουθίες.



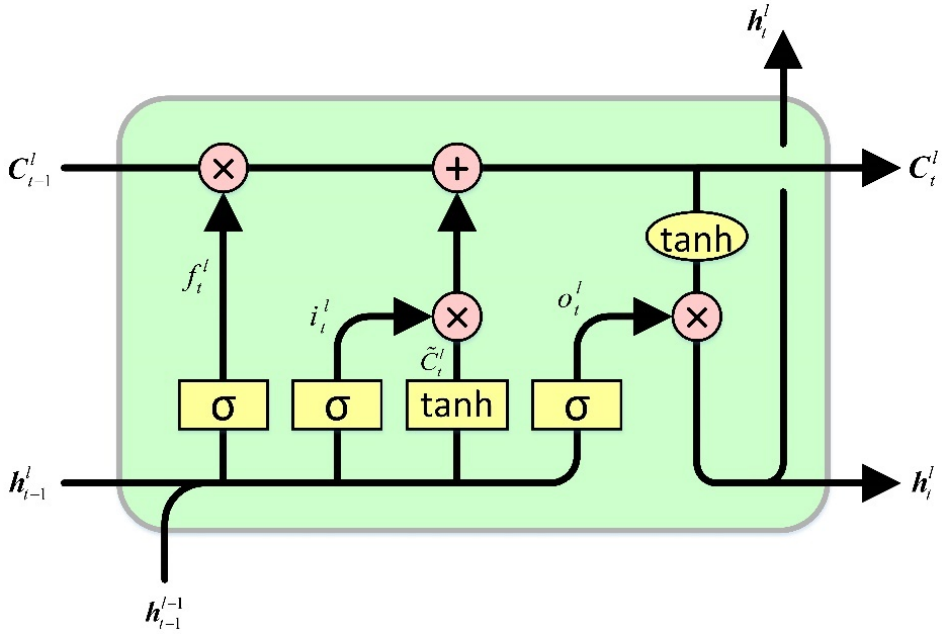
Εικόνα 24. Recurrent Neural Network

Πηγή: https://medium.com/analytics-vidhya/recurrent-neural-network-and-its-variants-de75f9ee063/

5.2.4 Long Short-Term Memory (LSTM)

Τα LSTM [48] έχουν σχεδιαστεί για να καταπολεμήσουν το πρόβλημα της εξαφάνισης ή της έκρηξης των gradient που συνδέεται με τα παραδοσιακά RNN ενσωματώνοντας μια πιο σύνθετη ανκαλούμενη μονάδα γνωστή ως cell (κύτταρο). Αυτά τα κύτταρα αποτελούνται από διάφορες πόρτες εισόδου, λήθης και εξόδου που ελέγχουν τη ροή των πληροφοριών εντός του κυττάρου, εξασφαλίζοντας ότι οι χρήσιμες πληροφορίες μένουν για μεγαλύτερες ακολουθίες. Αυτή η δομή επιτρέπει στα LSTM να συλλαμβάνουν τόσο τις μακροπρόθεσμες όσο και τις βραχυπρόθεσμες εξαρτήσεις στα δεδομένα, κάνοντας τα εξαιρετικά κατάλληλα για εργασίες όπως η πρόβλεψη χρονοσειρών, η επεξεργασία φυσικής γλώσσας και άλλοι τομείς.

Το αντίκτυπο που έχουν τα LSTM στους τομείς της μηχανικής μάθησης και της νευροϋπολογιστικής είναι τεράστιο. Έχουν βελτιώσει σημαντικά τις επιδόσεις σε ένα εύρος εφαρμογών, όπως η αναγνώριση ομιλίας, η μηχανική μετάφραση και άλλοι τομείς που απαιτούν τη μοντελοποίηση ακολουθιών. Για παράδειγμα, τα LSTM έχουν συμβάλει καθοριστικά στη βελτίωση της τεχνολογίας αναγνώρισης ομιλίας της Google και των μηχανικών μεταφράσεων στο Google Translate. Όπως επίσης από το Facebook για μεταφράσεις, αποδεικνύοντας την αποτελεσματικότητά τους σε πραγματικές εφαρμογές. Τα LSTM είναι γνωστά για την ικανότητά τους να μοντελοποιούν και να προβλέπουν μη γραμμικές δυναμικές συστημάτων που μεταβάλλονται στο χρόνο, καθιστώντας το έναν από τους πιο ισχυρούς δυναμικούς ταξινομητές που είναι γνωστοί μέχρι σήμερα.



Εικόνα 25. LSTM Cell

Πηγή: https://stackoverflow.com/questions/50488427/what-is-the-architecture-behind-the-keras-lstm-cell

6. Πειραματικό Μέρος

Σε αυτό το μέρος της πτυχιακής εργασίας περιγράφονται τα βήματα υλοποίησης του πειραματικού μέρους. Χρησιμοποιήθηκαν τρία διαφορετικά μοντέλα μηχανικής μάθησης με ένα σύνολο δεδομένων που προμηθεύτηκα από έμπιστες πηγές τους διαδικτύου. Τα μοντέλα εκπαίδευσης μέσω μηχανικής μάθησης που χρησιμοποίησα ήταν τα εξής:

- TCM (Temporal Convolutional Network)
- LSTM (Long Short-Term Memory)
- DNN (Deep Neural Network)

Παρακάτω θα εξηγηθούν οι λόγοι που διάλεξα αυτούς τους αλγορίθμους μαζί με μια μικρή περιγραφή της μεθοδολογίας του κάθε μοντέλου και την σειρά που έπραξα στο καθένα. Στην συνέχεια, θα εξηγήσω μερικά πράγματα περί του συνόλου δεδομένων και της προ-επεξεργασίας των δεδομένων που έκανα. Έπειτα θα πω για τις μετρήσεις που χρησιμοποίησα για να αξιολογήσω το μοντέλο. Και τέλος θα δούμε τα αποτελέσματα και θα τα σχολιάσουμε. Το σύνολο δεδομένων που χρησιμοποιήθηκε είναι ένα αρκετά βασικό στα περισσότερα σύνολα δεδομένων που χρησιμοποιούνται για αυτού του είδους έρευνας. Η διεύθυνση είναι η εξής 10.24342/f49465b2-c68a-4182-9171-075f0ed797d5 . Περιέχει δεδομένα χωρισμένα και με ετικέτες οι περισσότερες από τις οποίες διαγράφηκαν για καλύτερη εφαρμογή στα μοντέλα μου. Οι στήλες οι οποίες έμειναν μετά από αυτή την μικρή επεξεργασία είναι οι διεύθυνσης URL και ο τύπος της διεύθυνσης (Κακόβουλη ή όχι).

6.1 Πρώτο πείραμα

Στο πρώτο πείραμα χρησιμοποίησα ένα απλό και βασικό μοντέλο DNN μαζί με έναν συνδυασμό προ επεξεργασίας των μοντέλων μηχανικής μάθησης. Η διαδικασία ξεκινάει με την εισαγωγή των απαραίτητων βιβλιοθηκών, οι βιβλιοθήκες που χρησιμοποιήθηκαν είναι της TensorFlow και άλλες γενικές της Python όπως matplotlib και sklearn. Στη συνέχεια γίνεται η προ επεξεργασία του συνόλου δεδομένων, η οποία ξεκινάει με την διαγραφή των στηλών που δεν μας χρειάζονται από το έτοιμο σύνολο δεδομένων σε περίπτωση που δεν έχουμε συγκεντρώσει εμείς οι ίδιοι τα δεδομένα και έχουμε αυτό το κομμάτι έτοιμο. Γίνεται μια σειρά από εξαγωγές χαρακτηριστικών από τα δεδομένα όπως για παράδειγμα protocols, domain names, directory και άλλα. Παρακάτω παρατίθενται και σχετικές εικόνες από την προ επεξεργασία που έγινε.

	url	type	actual_url	protocol	host	
0	http://nobell.it/70ffb52d079109dca5664cce6f317...	1.0	nobell.it/70ffb52d079109dca5664cce6f317373782/...	http	nobell.it	/70ffb52d079109dca5664cce6f3
1	http://www.dghjdgf.com/paypal.co.uk/cycgi-bin/...	1.0	www.dghjdgf.com/paypal.co.uk/cycgi-bin/webscrc...	http	www.dghjdgf.com	/paypal.co.uk/cycgi-bin/webscrc
2	http://serviciosbys.com/paypal.cgi.bin.get-int...	1.0	serviciosbys.com/paypal.cgi.bin.get-into.herf....	http	serviciosbys.com	/paypal.cgi.bin.get-into.he
3	http://mail.printakid.com/www.online.americane...	1.0	mail.printakid.com/www.online.americanexpress....	http	mail.printakid.com	/www.online.americanex
4	http://thewhiskeydregs.com/wp-content/themes/w...	1.0	thewhiskeydregs.com/wp-content/themes/widescre...	http	thewhiskeydregs.com	/wp-content/themes/widescre

Εικόνα 26. Πρώτες επεξεργασίες των δεδομένων

Όπως θα δούμε και στην επόμενη εικόνα όλα τα χαρακτηριστικά θα καταλήξουν να είναι σε αριθμητικές τιμές.

	0	1	2	3	4	5	6	7	8
type	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
ip_address	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
URL__	4.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
URL_&	3.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
URL_~	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...
host_length	9.0	15.0	16.0	18.0	19.0	25.0	28.0	19.0	14.0
host_digits_in_url	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0
host_letters_in_url	8.0	13.0	15.0	16.0	18.0	23.0	27.0	18.0	9.0
host_vowels_url	3.0	1.0	5.0	6.0	5.0	8.0	9.0	5.0	4.0
protocol_http	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Εικόνα 27. Αριθμητικές τιμές των επεξεργασμένων δεδομένων

Στην συνέχεια όποια στήλη δεν χρειάζεται επειδή είναι NULL ή αχρείαστη γενικότερα, την αφαιρούμε. Αφού μπουν όλες οι ετικέτες που χρειαζόμαστε, χωρίζουμε το σύνολο δεδομένων σε Training Validation και Test Samples, Training είναι το δείγμα στο οποίο το μοντέλο θα εκπαιδευτεί, Validation είναι το κομμάτι που θα χρησιμοποιηθεί για να αξιολογηθεί ή να επικυρωθεί το μοντέλο και να δούμε το πως τρέχει σε δεδομένα που έχει ξαναδεί, Test είναι τα δεδομένα που το μοντέλο δεν έχει ξαναδεί και θα χρησιμοποιηθούν πρώτη φορά για την πραγματική δοκιμή.

Έπειτα έρχεται η εκπαίδευση του μοντέλου, έχουμε χρησιμοποιήσει μια διαδικασία που λέγεται k-fold cross-validation, μια τεχνική που αξιολογεί το πόσο καλά ένα μοντέλο θα λειτουργήσει σε ξεχωριστά δεδομένα. Το σύνολο δεδομένων χωρίζεται σε "k" αριθμό από ομοιόμορφα μέρη (folds). Στη συνέχεια, το μοντέλο εκπαιδεύεται "k" φορές, κάθε φορά με διαφορετικό fold ως σετ επικύρωσης και τα υπόλοιπα ως σετ εκπαίδευσης.

Η απόδοση του μοντέλου αξιολογείται κάθε φορά στο σετ επικύρωσης, παρέχοντας μια σταθερή εκτίμηση. Στα επίπεδα του μοντέλου όπως προαναφέραμε έχουμε χρησιμοποιήσει ένα απλό συνδυασμό επιπέδων για να εξετάσουμε πως λειτουργεί γενικότερα αυτός ο συνδυασμός τεχνικών μαζί με αυτό το σύνολο δεδομένων. Τα επίπεδα του μοντέλου μας 5 από 256 μέχρι 1 νευρώνα μιας και θέλουμε μια απάντηση για το εάν

το URL είναι κακόβουλο ή όχι. Τέλος φτιάχνουμε τα σχεδιαγράμματα τα οποία θα δούμε παρακάτω.

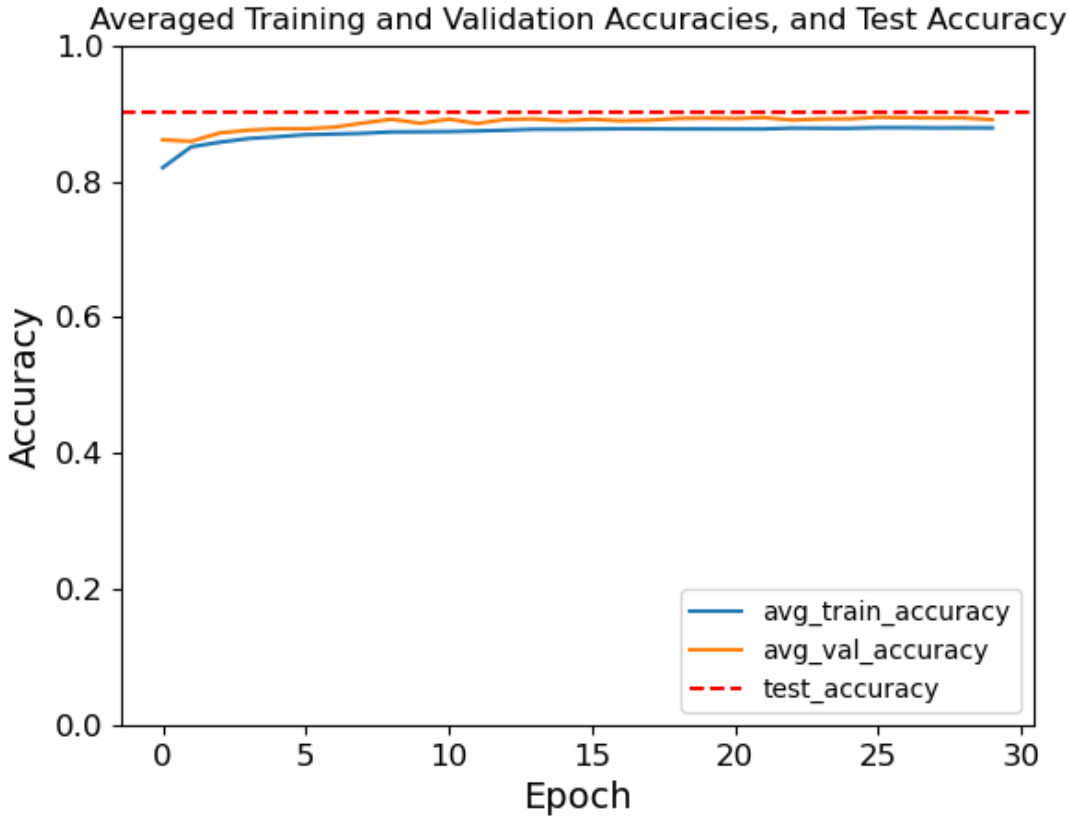
```
Test accuracy after retraining on trainval data: 0.9010
599/599 [==============================] - 1s 2ms/step
              precision    recall  f1-score   support

         0.0       0.88      0.93      0.90      9556
         1.0       0.92      0.87      0.90      9595

    accuracy                           0.90     19151
```

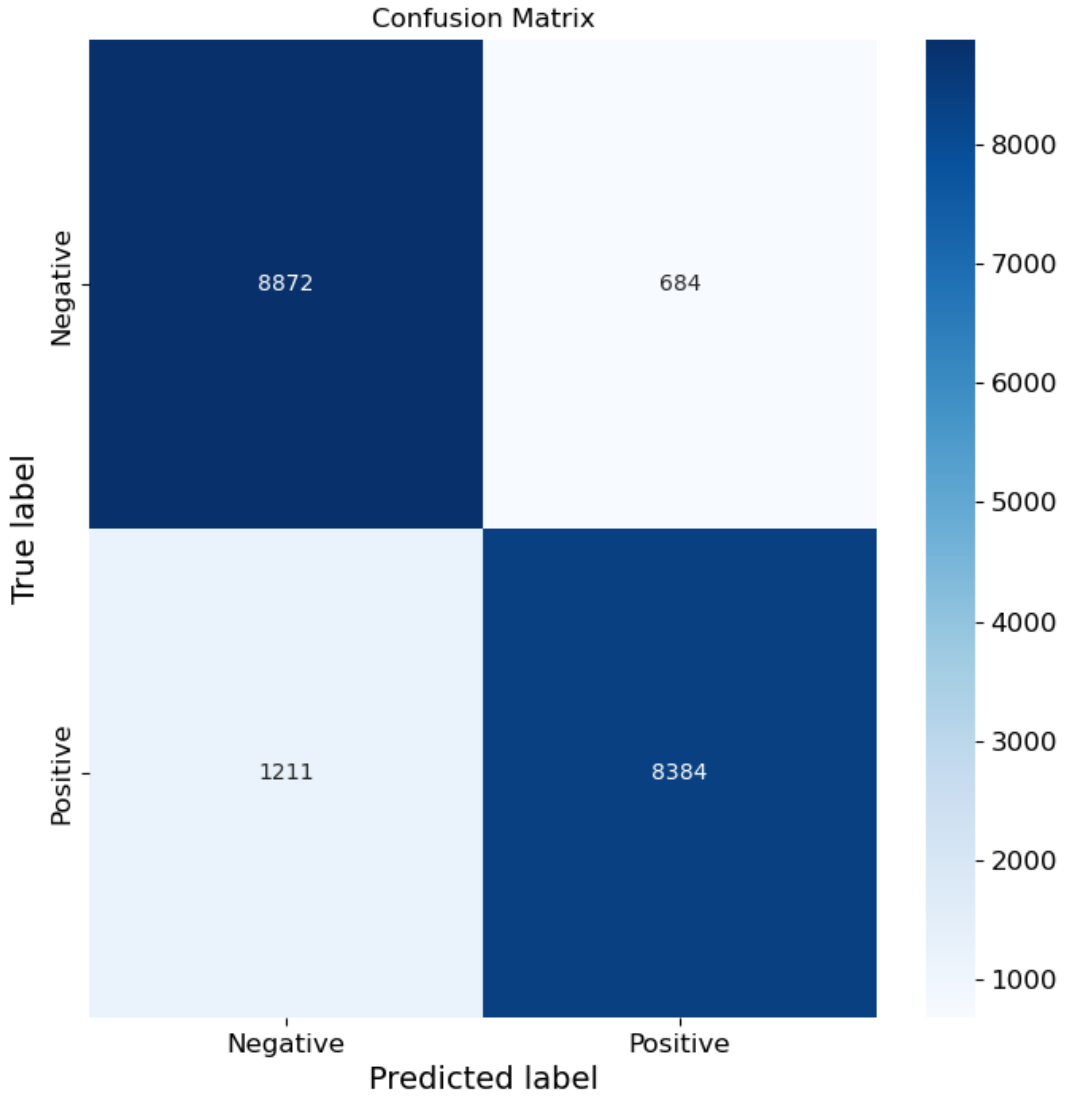
Εικόνα 28. DNN Output

Εδώ βλέπουμε precision, recall, f1-score και accuracy, παρακάτω θα τα δούμε και γραφικά.



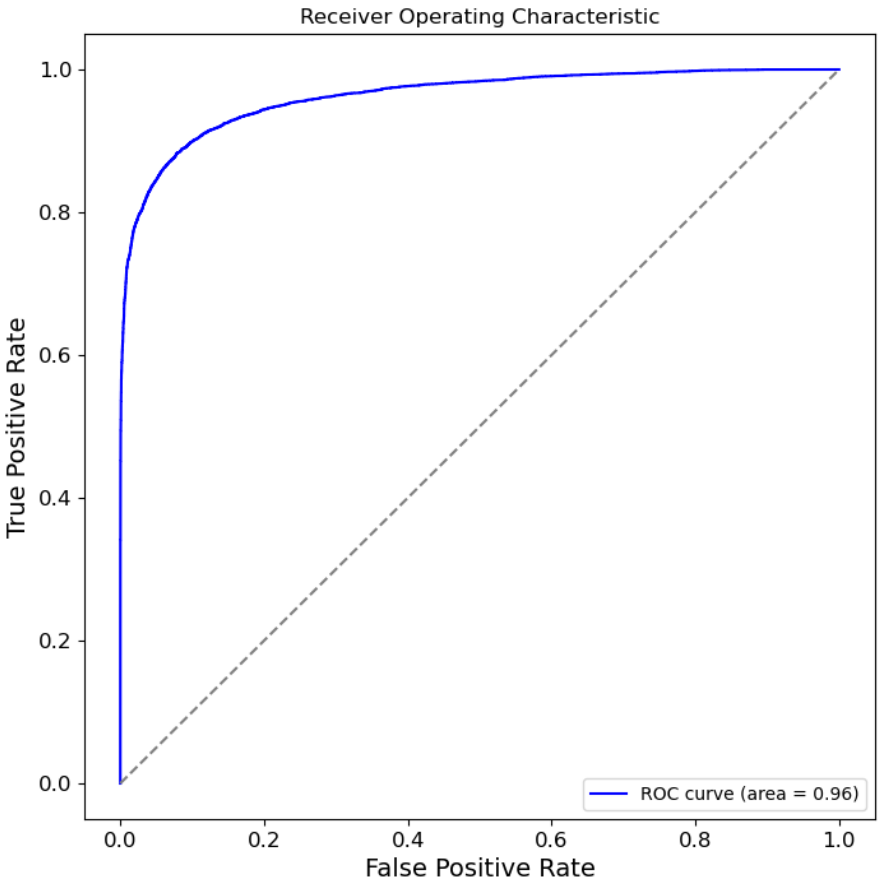
Εικόνα 29. Train, Val, Test DNN

Στον άξονα Χ έχουμε τις φορές που όλο το dataset έχει περάσει από την εκπαίδευση. Στον άξονα Υ παρουσιάζεται η ακρίβεια του μοντέλου που είναι ένα κομμάτι των σωστών προβλέψεων από όλες τις προβλέψεις. Και στα γραφήματα έχουμε την μέση ακρίβεια εκπαίδευσης , την μέση ακρίβεια αξιολόγησης και την ακρίβεια της δοκιμής. Βλέπουμε πως όσο περνάνε τα epochs τόσο καλύτερη γίνεται η ακρίβεια. Η ακρίβεια της δοκιμής είναι σταθερή γιατί γίνεται μια φορά στο τέλος.



Εικόνα 30. Confusion Matrix DNN

Έχουμε εξηγήσει προηγούμενος πως λειτουργεί το confusion matrix αλλά ας το αναλύσουμε λίγο παραπάνω. TN είναι 8872, TP είναι 8384, FN είναι 1211, FP είναι 684. Το μοντέλο αναγνώρισε σωστά τη μεγάλη πλειοψηφία τόσο των θετικών όσο και των αρνητικών δειγμάτων, όπως βλέπουμε στα True Positive και True Negative. Ο αριθμός των False Positive είναι σχετικά χαμηλός, γεγονός που υποδηλώνει ότι το μοντέλο δεν ταξινομεί συχνά λανθασμένα ένα αρνητικό δείγμα ως θετικό. Ο αριθμός των False Negative είναι επίσης σχετικά χαμηλός, αλλά υψηλότερος από τα False Positive. Αυτό σημαίνει ότι το μοντέλο ταξινομεί συχνότερα λανθασμένα ένα θετικό δείγμα ως αρνητικό σε σύγκριση με το αντίθετο. Σε τέτοια ζητήματα είναι σημαντικό τα False Negative να είναι όσο το δυνατόν λιγότερα



Εικόνα 31. Roc DNN

Η περιοχή κάτω από την καμπύλη ROC συμβολίζεται ως area = 0,96 σε αυτό το γράφημα. Η ROC (μπλε γραμμή) όπως είπαμε ξανά, όσο πιο μακριά και πάνω αριστερά είναι τόσο πιο αποδοτικό είναι το μοντέλο. Η γραμμή στο κέντρο υποδηλώνει μια τυχαία ταξινόμηση. Εδώ, με AUC 0,96, υποδηλώνει έναν πολύ καλό στην ταξινόμηση.

6.2 Δεύτερο πείραμα

Στο δεύτερο πείραμα χρησιμοποίησα τον αλγόριθμο TCN διότι ο αλγόριθμος αυτός φημίζεται για τον παραλληλισμό του, με λίγα λόγια μπορεί να δουλεύει με ολόκληρες ακολουθίες δεδομένων το οποίο βοηθάει στην ταχύτητα. Επίσης επειδή έχει ευέλικτα πεδία μπορεί να ανιχνεύει μοτίβο σε μεγάλες ακολουθίες, πέρα από αυτό η χρήση συνελίξεων είναι λιγότερο επιρρεπείς στα προβλήματα εξαφάνισης σε σύγκριση με τις άλλες επιλογές. Τέλος στην περίπτωση μου που δεν έχω την υπολογιστική ισχύ όσον αφορά την χωρητικότητα, τα TCN φημίζονται για την αποδοτικότητα που έχουν με την μνήμη.

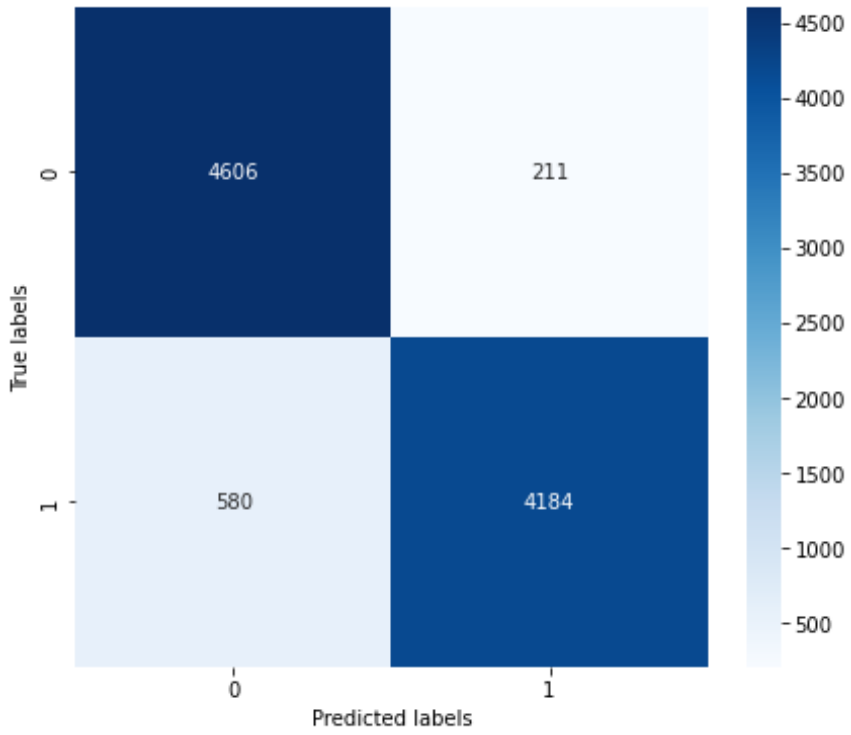
Η διαδικασία ξεκινάει με την εισαγωγή των απαραίτητων βιβλιοθηκών, οι βιβλιοθήκες που χρησιμοποιήθηκαν είναι κυρίως της Keras για το νευρωνικά δίκτυο, Pandas για την μεταχείριση των δεδομένων, sklearn για την διαχώριση των δεδομένων και matplotlib για τα γραφήματα. Στην συνέχεια αφού διαβάσουμε το σύνολο δεδομένων που έχουμε,

χρησιμοποιούμε έναν Tokenizer, είναι ένας τρόπος διαχωρισμού ενός κειμένου σε μικρότερες μονάδες που ονομάζονται tokens. Εδώ, οι μονάδες μπορεί να είναι είτε λέξεις, είτε χαρακτήρες, είτε κομμάτια λέξεων. Εγώ μετά από αρκετά trial and errors χρησιμοποιώ τα tokens σε επίπεδο χαρακτήρων, δηλαδή ένας χαρακτήρας ένα token. Τα tokens χρειάζονται για να μετατρέπονται τα γράμματα σε ακολουθίες που μπορούν να ερμηνευθούν από τον αλγόριθμο. Έπειτα γίνεται το padding το οποίο βοηθάει στο να έχουν όλες οι ακολουθίες το ίδιο μήκος, έτσι γεμίζει με μηδενικά τα κενά από όλα τα URL που είναι μικρότερα από το μεγαλύτερο. Συνεχίζοντας χρησιμοποιώ την στήλη type για ετικέτα. Χωρίζω τα δεδομένα σε Training , Validation και Test δείγματα και ύστερα τα ξ ανασχηματίζουμε για να μπορούν να χωρέσουν στο μοντέλο. Τώρα, χτίζουμε το μοντέλο μας με νευρωνικά δίκτυα με επίπεδα convolutional 1D δηλαδή με συνελικτικά μιας διάστασης και με το dilation βοηθάμε το μοντέλο στην γενίκευση όταν αλλάζει από το ένα επίπεδο στο άλλο. Τέλος γίνεται η εκπαίδευση και τα γραφήματα με τις μετρήσεις.

	precision	recall	f1-score	support
0.0	0.89	0.96	0.92	4817
1.0	0.95	0.88	0.91	4764
accuracy			0.92	9581

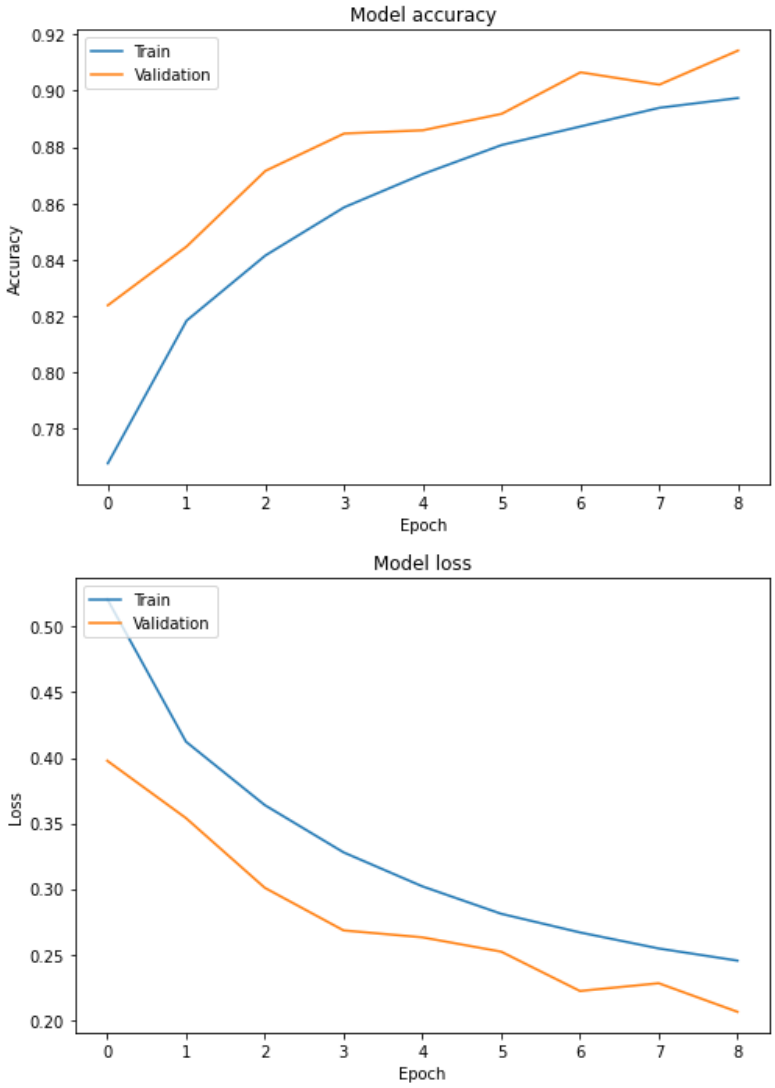
Εικόνα 32. TCN Output

Εδώ βλέπουμε αναλυτικά τα precision, recall, f1-score. Όπως και στο προηγούμενο μοντέλο έτσι και εδώ είναι η ίδια διάταξη, με καλύτερα όμως αποτελέσματα. 0.0 αντιστοιχεί στα Malicious και το 1.0 στα Benign.



Εικόνα 33. TCN Confusion Matrix

TN είναι 4606, TP είναι 4184, FN είναι 580, FP είναι 211. Το μοντέλο αναγνώρισε σωστά τη μεγάλη πλειοψηφία τόσο των θετικών όσο και των αρνητικών δειγμάτων, όπως βλέπουμε στα True Positive και True Negative. Ο αριθμός των False Positive είναι σχετικά χαμηλός, γεγονός που υποδηλώνει ότι το μοντέλο δεν ταξινομεί συχνά λανθασμένα ένα αρνητικό δείγμα ως θετικό. Ο αριθμός των False Negative είναι επίσης σχετικά χαμηλός, αλλά υψηλότερος από τα False Positive. Αυτό σημαίνει ότι το μοντέλο ταξινομεί συχνότερα λανθασμένα ένα θετικό δείγμα ως αρνητικό σε σύγκριση με το αντίθετο. Σε τέτοια ζητήματα είναι σημαντικό τα False Negative να είναι όσο το δυνατόν λιγότερα όπως αναφέραμε και στο προηγούμενο πείραμα.



Εικόνα 34. TCN Accuracy & Loss

Στον άξονα Χ έχουμε τις φορές που όλο το dataset έχει περάσει από την εκπαίδευση. Στον άξονα Υ παρουσιάζεται η ακρίβεια του μοντέλου που είναι ένα κομμάτι των σωστών προβλέψεων από όλες τις προβλέψεις και το πόσα δεδομένα έχουν χαθεί στο δεύτερο γράφημα. Και στα γραφήματα έχουμε την μέση ακρίβεια εκπαίδευσης , την μέση ακρίβεια αξιολόγησης. Βλέπουμε πως όσο περνάνε τα epochs τόσο καλύτερη γίνεται η ακρίβεια και τόσο μειώνεται το loss.

6.3 Τρίτο πείραμα

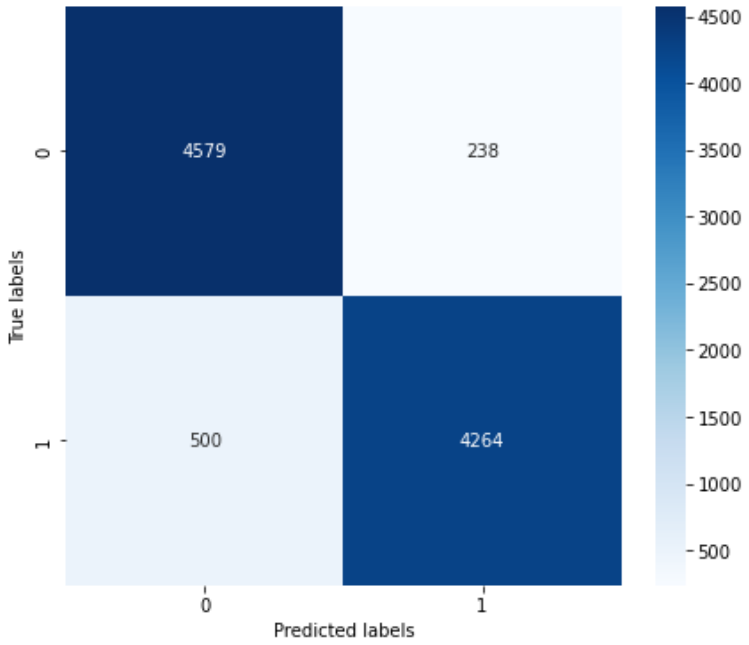
Στο τρίτο πείραμα χρησιμοποίησα τον αλγόριθμο LSTM διότι ο αλγόριθμος αυτός φημίζεται για την διαχείριση ακολουθιών και την ικανότητά του στην καταγραφή μακροπρόθεσμων εξαρτήσεων. Αυτό είναι πολύ χρήσιμο για την ανίχνευση URL, μιας και η φορές που θα εμφανιστεί κάθε χαρακτήρας μπορεί να εξαρτάται από χαρακτήρες που εμφανίστηκαν πολύ νωρίτερα στην ακολουθία. Επίσης μπορεί να δοθεί ιδιαίτερη έμφαση στα σημαντικά κομμάτια των URL με την χρήση των πυλαίων των LSTM που κρατάνε την

πιο σημαντική πληροφορία. Όλη η διαδικασία μέχρι την χρήση του μοντέλου είναι η ίδια με αυτή του δεύτερου πειράματος. Σε αυτό το μοντέλο χρησιμοποιούμε 4 επίπεδα 32 64 και 32 κόμβων που μεταφέρουν όλοί την ακολουθία και στο τέλος ένα dense layer με sigmoid για να βγάλει την πιθανότητα του 0 ή του 1. Στην συνέχεια παραθέτω τα αποτελέσματα του training.

	precision	recall	f1-score	support
0.0	0.90	0.95	0.93	4817
1.0	0.95	0.90	0.92	4764
accuracy			0.92	9581

Εικόνα 35. LSTM Output

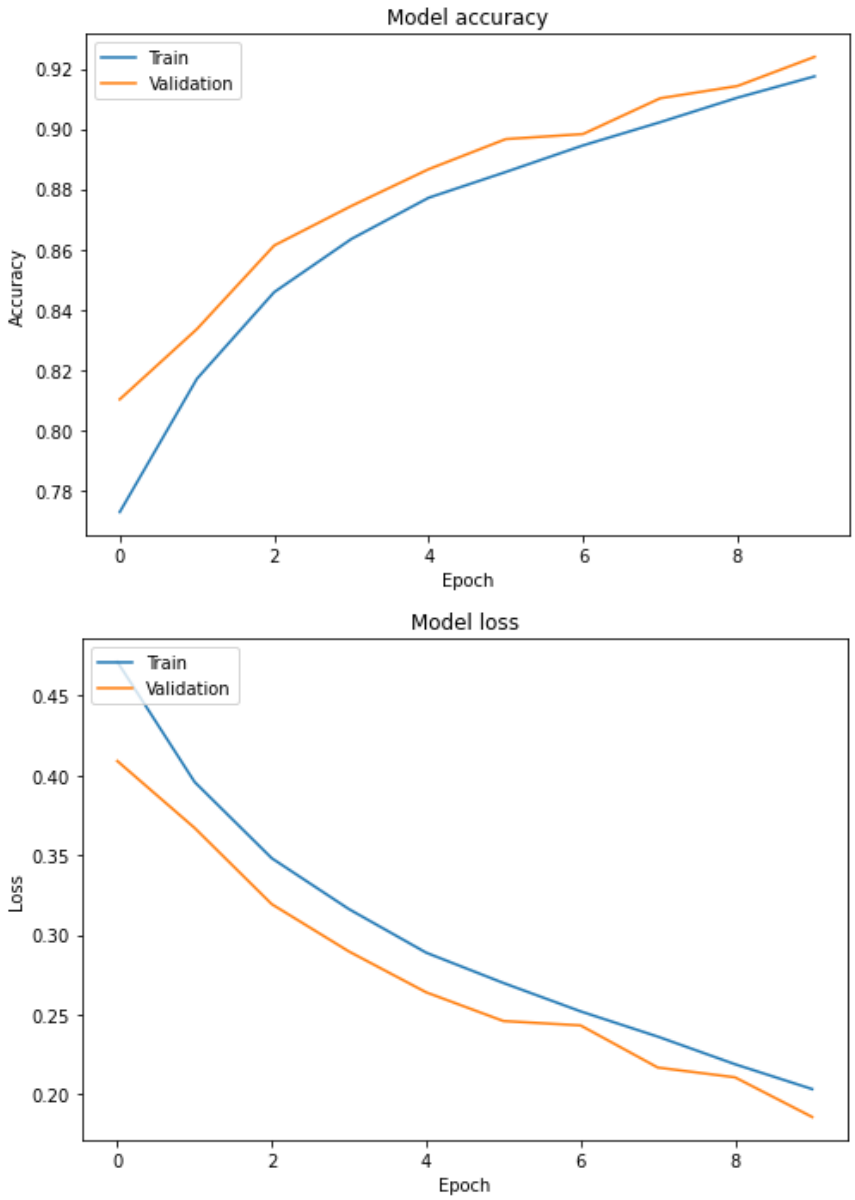
Εδώ βλέπουμε αναλυτικά τα precision, recall, f1-score. Όπως και στο προηγούμενο μοντέλο έτσι και εδώ είναι η ίδια διάταξη. 0.0 αντιστοιχεί στα Malicious και το 1.0 στα Benign.



Εικόνα 36. LSTM Confusion Matrix

TN είναι 4579, TP είναι 4264, FN είναι 500, FP είναι 238. Το μοντέλο αναγνώρισε σωστά τη μεγάλη πλειοψηφία τόσο των θετικών όσο και των αρνητικών δειγμάτων, όπως βλέπουμε στα True Positive και True Negative. Ο αριθμός των False Positive είναι σχετικά χαμηλός, γεγονός που υποδηλώνει ότι το μοντέλο δεν ταξινομεί συχνά λανθασμένα ένα αρνητικό δείγμα ως θετικό. Ο αριθμός των False Negative είναι επίσης σχετικά χαμηλός, αλλά υψηλότερος από τα False Positive. Αυτό σημαίνει ότι το μοντέλο ταξινομεί συχνότερα λανθασμένα ένα θετικό δείγμα ως αρνητικό σε σύγκριση με το αντίθετο. Σε τέτοια

ζητήματα είναι σημαντικό τα False Negative να είναι όσο το δυνατόν λιγότερα όπως αναφέραμε και στο προηγούμενο πείραμα.



Εικόνα 37. LSTM Accuracy & Loss

Στον άξονα Χ έχουμε τις φορές που όλο το dataset έχει περάσει από την εκπαίδευση. Στον άξονα Υ παρουσιάζεται η ακρίβεια του μοντέλου που είναι ένα κομμάτι των σωστών προβλέψεων από όλες τις προβλέψεις και το πόσα δεδομένα έχουν χαθεί στο δεύτερο γράφημα. Και στα γραφήματα έχουμε την μέση ακρίβεια εκπαίδευσης , την μέση ακρίβεια αξιολόγησης. Βλέπουμε πως όσο περνάνε τα epochs τόσο καλύτερη γίνεται η ακρίβεια και τόσο μειώνεται το loss.

6.4 Συγκριτικά Αποτελέσματα Πειραμάτων

Στον Πίνακα 2 αποτυπώνονται τα αποτελέσματα από την υλοποίηση των τριών πειραμάτων με την εφαρμογή των επιλεγμένων μοντέλων εκπαίδευσης μέσω μηχανικής μάθησης (DNN, TCN, LSTM):

Πίνακας 2. Αποτελέσματα Μοντέλων Εκπαίδευσης μέσω Μηχανικής Μάθησης

Metrics	DNN	TCN	LSTM
Accuracy	0.9010673928492533	0.9174407720565796	0.9229725599288941
Precision	0.9011389589487298	0.9519909024238586	0.9471346139907837
Recall	0.8976473836320266	0.8782535791397095	0.8950461745262146
Loss	0.2127483946456323	0.2022259533405304	0.1893843561410904
AUC	0.9643546392048264	0.9737770557403564	0.9772130846977234

Με βάση τα αποτελέσματα των τριών μοντέλων (DNN, TCN, LSTM) που χρησιμοποιήθηκαν για την ταξινόμηση των διευθύνσεων URL ως καλοήθεις ή κακόβουλες, μπορεί να εξαχθεί το συμπέρασμα ότι:

- **Accuracy**: Το μοντέλο LSTM έχει την υψηλότερη ακρίβεια 0,923, ακολουθούμενο από το μοντέλο TCN με ακρίβεια 0,917 και το μοντέλο DNN με ακρίβεια 0,901. Αυτό δείχνει ότι το μοντέλο LSTM είναι το πιο ακριβές στην σωστή ταξινόμηση των URL ως καλοήθεις ή κακόβουλες.
- **Precision**: Το μοντέλο TCN έχει την υψηλότερη ορθότητα 0,952, λίγο περισσότερο από το μοντέλο LSTM με ορθότητα 0,947 και το μοντέλο DNN με ορθότητα 0,901. Αυτό υποδηλώνει ότι το μοντέλο TCN είναι το καλύτερο στην ελαχιστοποίηση των ψευδώς θετικών αποτελεσμάτων.
- **Recall**: Το μοντέλο LSTM έχει την υψηλότερη ανάκληση 0,895, ακολουθούμενο από το μοντέλο DNN με ανάκληση 0,898 και το μοντέλο TCN με ανάκληση 0,878. Αυτό σημαίνει ότι το μοντέλο LSTM είναι το καλύτερο στην ελαχιστοποίηση των ψευδώς αρνητικών αποτελεσμάτων.
- **Loss**: Το μοντέλο LSTM έχει τη χαμηλότερη απώλεια 0,189, ακολουθούμενο από το μοντέλο TCN με απώλεια 0,202 και το μοντέλο DNN με απώλεια 0,213. Αυτό δείχνει ότι το μοντέλο LSTM έχει το μικρότερο ποσό σφάλματος στις προβλέψεις του.
- **AUC**: Το μοντέλο LSTM έχει την υψηλότερη AUC 0,977, ακολουθούμενο από το μοντέλο TCN με AUC 0,974 και το μοντέλο DNN με AUC 0,964. Αυτό υποδηλώνει

ότι το μοντέλο LSTM είναι το καλύτερο στη διάκριση μεταξύ καλοηθών και κακόβουλων διευθύνσεων URL.

Συμπερασματικά, ενώ και τα τρία μοντέλα αποδίδουν καλά στην ταξινόμηση των URL ως καλοήθεις ή κακόβουλες, το μοντέλο LSTM υπερέχει των άλλων δύο μοντέλων στις περισσότερες μετρήσεις, καθιστώντας το, το καταλληλότερο μοντέλο για αυτή την εργασία. Ωστόσο, η ανώτερη ακρίβεια του μοντέλου TCN υποδεικνύει τη δυνητική χρησιμότητά του σε σενάρια όπου η ελαχιστοποίηση των ψευδώς θετικών αποτελεσμάτων είναι ιδιαίτερα κρίσιμη. Επομένως, η επιλογή του μοντέλου μπορεί να εξαρτάται από τις συγκεκριμένες απαιτήσεις και τους περιορισμούς της εφαρμογής.

Συμπεράσματα

Σκοπός της παρούσας εργασίας είναι η βελτίωση της μεθοδολογίας για την ανίχνευση σε πραγματικό χρόνο κακόβουλων επιθέσεων ηλεκτρονικού ψαρέματος (phishing) με τη χρήση τεχνολογίας μηχανικής μάθησης. Ο στόχος μας είναι να αποτρέψουμε αυτούς τους χρήστες που σερφάρουν στο διαδίκτυο από το να χάσουν τα προσωπικά τους δεδομένα, τραπεζικά στοιχεία, αρχεία, έγγραφα, ψηφιακά νομίσματα από το phishing. Για την καλύτερη πρόβλεψη της ιστοσελίδας phishing μπορούν να χρησιμοποιηθούν διαφορετικοί αλγόριθμοι και μοντέλα μηχανικής μάθησης, που μπορούν να βοηθούν τους χρήστες να διακρίνουν αν ο δικτυακός τόπος που επισκέπτονται είναι ύποπτος για επίθεση phishing ή νόμιμος. Οι χρήστες χρησιμοποιούν τόσες πολλές τεχνικές για να αποτρέψουν τις επιθέσεις, αν και δεν είναι πάντοτε επιτυχείς.

Στα πλαίσια του πειραματικού μέρους της εργασίας εξετάζουμε την αξιοποίηση τριών μοντέλων εκπαίδευσης (DNN, TCN, LSTM) μέσω μηχανικής μάθησης, για να ανιχνεύουμε σε πραγματικό χρόνο κακόβουλες επιθέσεις ηλεκτρονικού ψαρέματος, μέσω ψευδεπίγραφων-πλαστών διευθύνσεων URL και να τις διακρίνουμε από τις νόμιμες διευθύνσεις URL. Έχουμε και δοκιμής για να ανιχνεύσουμε ότι η δεδομένη διεύθυνση URL είναι εξουσιοδοτημένη (καλοήθης) ή πλαστή (κακόβουλη). Μετά την εξαγωγή των χαρακτηριστικών εφαρμόζουμε τις διαφορετικές λειτουργίες εκπαίδευσης και δοκιμής για να λάβουμε τα αποτελέσματα (δηλαδή να προβλέψουμε αν η συγκεκριμένη διεύθυνση URL είναι είτε πραγματική είτε κακόβουλη). Βάσει αυτού προσπαθήσαμε να πάρουμε τα αποτελέσματα του κάθε αλγορίθμου, να βρούμε το ποσοστό ακρίβειας και να συγκρίνουμε μεταξύ τους προκειμένου να επιτύχουμε την υψηλότερη ακρίβεια και να χρησιμοποιήσουμε αυτούς τους αλγορίθμους για την πρόβλεψη καλύτερου αποτελέσματος με την υψηλότερη ακρίβεια.

Προτάσεις Μελλοντικής Επέκτασης

Ως μελλοντική επέκταση της παρούσας εργασίας, μπορούν να εξεταστούν διάφοροι τρόποι για τη βελτίωση και την επέκταση των δυνατοτήτων των μοντέλων ταξινόμησης URL. Κατ' αρχάς, θα μπορούσε να διερευνηθούν μέθοδοι συνόλου, που συνδυάζουν τα πλεονεκτήματα πολλαπλών μοντέλων, ώστε να βελτιωθεί ενδεχομένως η συνολική απόδοση. Τεχνικές συνόλου, όπως η stacking ή bagging, μπορεί να παρέχουν μια πιο ισχυρή και γενικευμένη λύση. Επιπλέον, θα μπορούσε να πραγματοποιηθεί περαιτέρω hyperparameter tuning για τη λεπτομερή ρύθμιση των υφιστάμενων μοντέλων, εξασφαλίζοντας τη βέλτιστη απόδοση σε συγκεκριμένα σύνολα δεδομένων ή υπό διαφορετικές συνθήκες.

Η διερεύνηση της ενσωμάτωσης πρόσθετων features θα μπορούσε να συμβάλει σε μια πιο ολοκληρωμένη κατανόηση των χαρακτηριστικών των διευθύνσεων URL, βελτιώνοντας

ενδεχομένως την ικανότητα των μοντέλων να διακρίνουν μεταξύ κακόβουλων και καλοηθών διευθύνσεων URL. Το feature engineering ή η ενσωμάτωση εξωτερικών πηγών δεδομένων μπορεί να προσφέρει πολύτιμες πληροφορίες για τα αναδυόμενα πρότυπα απειλών.

Επιπλέον, δεδομένης της εξελισσόμενης φύσης των απειλών στον κυβερνοχώρο, η συνεχής παρακολούθηση και ενημέρωση των μοντέλων είναι ζωτικής σημασίας. Η ανάπτυξη μηχανισμών για την προσαρμογή του μοντέλου σε δυναμικά μεταβαλλόμενα τοπία διευθύνσεων URL, ενδεχομένως αξιοποιώντας προσεγγίσεις διαδικτυακής μάθησης, θα μπορούσε να αποτελέσει μια πολλά υποσχόμενη κατεύθυνση για μελλοντική έρευνα.

Τέλος, η ανάπτυξη και η επικύρωση των μοντέλων σε πραγματικό περιβάλλον θα αποτελούσε σημαντικό βήμα προς την αξιολόγηση της πρακτικής χρησιμότητάς τους. Αυτό περιλαμβάνει τη δοκιμή των μοντέλων σε ποικίλα και εξελισσόμενα σύνολα δεδομένων και την εξέταση του αντίκτυπου των ψευδώς θετικών και ψευδώς αρνητικών αποτελεσμάτων σε πραγματικές επιχειρήσεις ασφαλείας. Η αντιμετώπιση των προβλημάτων κλιμάκωσης και υπολογιστικής αποδοτικότητας για την ανάπτυξη μεγάλης κλίμακας αποτελεί επίσης ένα σχετικό θέμα για τις μελλοντικές ερευνητικές προσπάθειες.

Συνοψίζοντας, οι μελλοντικές εργασίες στην ταξινόμηση URL θα μπορούσαν να επικεντρωθούν σε μεθόδους συνόλου, εμπλουτισμό χαρακτηριστικών, προσαρμογή του μοντέλου στις εξελισσόμενες απειλές και εκτιμήσεις για την ανάπτυξη σε πραγματικό κόσμο, ώστε να προωθηθεί περαιτέρω η αποτελεσματικότητα και η δυνατότητα εφαρμογής των μοντέλων που αναπτύχθηκαν.

Βιβλιογραφία

[1] Y. Li and Q. Liu, "A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments,," *Energy Reports,* vol. 7, pp. 8176-8186, 2021.

[2] A. Mohammed Saeed, C. Sajjad Hussain and A. A. Mohammed, "Seamless security apprise method for improving the reliability of sustainable energy-based smart home applications," *Sustainable Energy Technologies and Assessments,* p. volume 45, 2021.

[3] Obi Ogbanufe, «Enhancing End-User Roles in Information Security: Exploring the Setting, Situation, and Identity,,» *ScienceDirect,* p. Volume 108, 2021.

[4] K. Vidhyanandhini και V. Saravanarajan, «An efficient data flow material model based cloud authentication data security and reduce a cloud storage cost using Index-level Boundary Pattern Convergent Encryption algorithm,» *Materials Today: Proceedings,* τόμ. 81, αρ. 2, pp. 931-936, 2021.

[5] L. Qinghui and Z. Tianping, "Deep learning technology of computer network security detection based on artificial intelligence," *ScienceDirect,* vol. Volume 110, 2023.

[6] S. Alam, «Cybersecurity: Past, Present and Future,» Department of Computer Engineering, Adana Science and Technology University, 04 July 2022. [Ηλεκτρονικό]. Available: https://arxiv.org/ftp/arxiv/papers/2207/2207.01227.pdf. [Πρόσβαση 30 October 2023].

[7] T. R. Franklin, N. N. Corneille, V. C. Kamla και K. P. Udagepola, «LimonDroid: a system coupling three signature-based schemes for profiling Android malware,» *Iran Journal of Computer Science,* τόμ. 4, pp. 95-114, June 2021.

[8] C. C. Ahmet και B. K. Turkan, «The current state and future of mobile security in the light of the recent mobile security threat reports,» *Multimedia Tools and Applications,* τόμ. 82, p. 20269–20281, 30 January 2023.

[9] C. Zhenxiang, Y. Qiben, H. Hongbo, W. Shanshan, L. P. W. Lin και Y. Bo, «Machine learning based mobile malware detection using highly imbalanced network traffic,» τόμ. 433–434, pp. 346-364, April 2018.

[10] D. R. Gabi και G. Mario, «Cyber Security: Challenges and Application Areas,» σε *Supply Chain Safety Management*, Springer Link, 2022, pp. 179-197.

[11] T. Kutub, Q. Meikang, G. Keke και L. A. Md, «An Investigation on Cyber Security Threats and Security Models,» *IEEE Xplore,* 7 January 2016.

[12] S. A. Wasyihun, M. Yirga Yayeh και D. Abebe Abeshu, «Cyber security: State of the art, challenges and future directions,» *Cyber Security and Applications,* αρ. 2, 1 October 2023.

[13] H. Mamoona, N. Mahmood, N. J. A. Mohammad και M. Sajjad, «Cyber Security Threats and Vulnerabilities: A Systematic Mapping Study,» *Arabian Journal for Science and Engineering,* τόμ. 45, 6 January 2020.

[14] H. Nelson, «National Institute of Standards and Technology,» NIST, 8 July 2023. [Ηλεκτρονικό]. Available: https://www.nist.gov/programs-projects/cybersecurity-smart-grid-systems. [Πρόσβαση 30 October 2023].

[15] NHTSA, «Vehicle Cybersecurity,» NHTSA, 7 July 2022. [Ηλεκτρονικό]. Available: https://www.nhtsa.gov/technology-innovation/vehicle-cybersecurity/. [Πρόσβαση 30 October 2023].

[16] National Cyber Security Centre (CISA), «Cybersecurity Best Practices for Smart Cities,» 19 April 2023. [Ηλεκτρονικό]. Available: https://www.cisa.gov/sites/default/files/2023-04/cybersecurity-best-practices-for-smart-cities_508.pdf. [Πρόσβαση 31 October 2023].

[17] IBM, «IBM,» 2023. [Ηλεκτρονικό]. Available: https://www.ibm.com/topics/machine-learning. [Πρόσβαση 31 October 2023].

[18] B. Farhad, N. Jessie, N. Adel και Z. Tian, «Machine Learning Principles,» *Springer Link,* pp. 115-157, 23 January 2020.

[19] S. Kamran και A. H. Ibrahim, «A Survey on Machine Learning Techniques for,» *IEEE Access,* pp. 9-10, 2 December 2020.

[20] B. Aayush, «neptune.ai,» Neptune Labs, 29 September 2023. [Ηλεκτρονικό]. Available: https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide. [Πρόσβαση 1 November 2023].

[21] Google, «Google Developers,» 18 July 2022. [Ηλεκτρονικό]. Available: https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall. [Πρόσβαση 1 November 2023].

[22] M. Gabor, «Gabor Melli's Research Knowledge Base,» 4 October 2023. [Ηλεκτρονικό]. Available: https://www.gabormelli.com/RKB/Fallout_Precision. [Πρόσβαση 1

November 2023].

[23] P. Roel, «roelpeters.be,» 2023. [Ηλεκτρονικό]. Available: https://www.roelpeters.be/glossary/false-omission-rate/. [Πρόσβαση 1 November 2023].

[24] Google Developers, «developers.google,» Google, 18 July 2022. [Ηλεκτρονικό]. Available: https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc. [Πρόσβαση 1 November 2023].

[25] S. Allright, «Stephen Allright,» 2023. [Ηλεκτρονικό]. Available: https://stephenallwright.com/good-mse-value/. [Πρόσβαση 1 November 2023].

[26] G. Stephanie, «Statistics How To,» 2023. [Ηλεκτρονικό]. Available: https://www.statisticshowto.com/absolute-error/. [Πρόσβαση 1 November 2023].

[27] ZACH, «Statology,» Statology, 2023. [Ηλεκτρονικό]. Available: https://www.statology.org/mape-excel/. [Πρόσβαση 1 November 2023].

[28] N. Dewang, «ML| Underfitting and Overfitting,» 31 August 2023. [Ηλεκτρονικό]. Available: https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/. [Πρόσβαση 1 November 2023].

[29] «TechTarget,» 2023. [Ηλεκτρονικό]. Available: https://www.techtarget.com/searchenterpriseai/definition/supervised-learning. [Πρόσβαση 1 November 2023].

[30] M. Sanatan, «TowardsDataScience,» 20 May 2023. [Ηλεκτρονικό]. Available: https://towardsdatascience.com/unsupervised-learning-and-data-clustering-eeecb78b422a. [Πρόσβαση 1 November 2023].

[31] G. Aldo, E. H. Brooke, R. Alex, C. Cecilia, N. Frank και L. Alessandro, «Unsupervised Learning Methods for Molecular Simulation Data,» *Chemical Reviews,* p. 121, 4 May 2021.

[32] T. K. Tala, O. S. Hadjar και K. Naima, «Deep learning: systematic review, models, challenges, and research directions,» *ResearchGate,* September 2023.

[33] D. Dema, I. Fabiha, A. Zulfikar, A. Zeyar και A. A. Mohammad, «Reinforcement Learning: A Friendly Introduction,» *Springer Link,* pp. 134-146, 8 August 2021.

[34] S. Nimish, «Introduction to Reinforcement Learning,» *Deep Reinforcement Learning with Python,* pp. 1-17, 2 April 2021.

[35] L. Yann, B. Yoshua και H. Geoffrey, «Deep Learning,» *Research Gate,* May 2015.

[36] K. Khushbu και Y. Suniti, «Linear Regression Analysis Study,» *ResearchGate,* 4 May 2018.

[37] M. Maher, «Logistic Regression in Data Analysis: AnOverview,» *ResearchGate,* pp. 281-299, July 2011.

[38] S. Kotsiantis, «Decision trees: a recent overview,» *Artificial Intelligence Review,* αρ. 39, pp. 261-283, 29 June 2011.

[39] W. Geoffrey, «Naive Bayes,» *Encyclopedia of Machine Learning and Data Mining,* pp. 1-2, January 2016.

[40] L. Breiman, «Random Forests,» *Machine Learnng,* αρ. 45, pp. 5-32, October 2001.

[41] C. Padraig και D. Sarah Jane, «k-Nearest Neighbor Classifiers,» *ResearchGate,* April 2007.

[42] I. Sakshi, K. G. Anil, M. S.P. και A. Pooja, «Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach,» *Procedia Computer Science,* αρ. 132, pp. 679-688, 2018.

[43] O. Keiron και N. Ryan, «An Introduction to Convolutional Neural Networks,» *ResearchGate,* November 2015.

[44] Z. Ziyuan, W. Jianzhou, W. Danxiang και X. Yurui, «An improved temporal convolutional network with attention mechanism for photovoltaic generation forecasting,» *Research Gate,* August 2023.

[45] M. Hyangsuk και J.-G. L. , «Temporal Convolutional Network-Based,» *IEEE,* pp. 269-276, 2023 IEEE International Conference on Big Data and Smart Computing (BigComp).

[46] M. S. Robin, «Recurrent Neural Networks (RNNs): A gentle Introduction and Overview,» *arXiv,* 23 November 2019.

[47] G. Tarun Kumar και R. Khalid, «Recurrent Neural Network,» *ScienceDirect,* 2022.

[48] C. S. Ralf και R. M. Eric, «Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks,» 12 September 2019.